

**Digital privacy and new media: An empirical study
assessing the impact of Privacy Seals on personal
information disclosure.**

Volume 1 of 1

Conor Paul O'Kane

PhD

Bournemouth University

April 2019

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Conor Paul O'Kane

Digital privacy and new media: An empirical study assessing the impact of Privacy Seals on personal information disclosure.

Abstract

Advances in technology have facilitated the rapid growth of a global new media industry. Many new media firms rely heavily on networked technologies to enable a primary income driver based on advertising revenues. This has attracted criticisms from privacy campaigners who argue that elements of the way some of these firms operate constitute an invasion of user's privacy.

Early economic approaches to privacy are primarily informed by the rational choice theory and viewed individuals as utility maximizers when making decisions involving personal information disclosure. Theoretical approaches have since developed to account for factors explored by bounded rationality and behavioural economics where individuals engage in complex trade-offs when making privacy disclosure decisions.

Both EU and US regulators believe rapid technological advances have rendered existing regulatory provisions inadequate. In the EU, the 2018 General Data Protection Regulation (GDPR) set out to improve 'information transparency' and give individuals to exercise greater 'control' over their personal data. The regulation set out provisions for the establishment of a privacy seal accreditation scheme.

There is little empirical evidence to demonstrate that the use of privacy seals is privacy enhancing. Existing research reveals inconsistent and at times counter-intuitive findings. This research conducted online experimental research to establish if a causal link exists between the presence of a privacy seals and personal information disclose. Experiment results show that contrary to previous research in this area, the presence of privacy seals does not result in lower personal information disclosure. Survey findings also show that the GDPR has failed to expand 'sensitive' categories of data in line with both EU and US data subjects expectations.

This research makes a number of original contributions to knowledge. Information disclosure is examined in relation to sensitive data categories as defined in the GDPR. Using commercially available privacy seals, it adds to the existing body of literature on the impact of iconography on user behaviour. The findings suggest there is an opportunity for new media firms to use independently accredited privacy seals as a differentiator in this industry sector.

Table Of Contents

Abstract	3
1 Chapter 1: Introduction	11
1.1 New Media Privacy Issues in Context.....	12
1.2 The Rapid Emergence of 'New Media Industry'	15
1.3 New Media: A Culture of Disclosure?	18
1.4 Regulating Privacy.....	22
1.5 Research Gap	24
1.6 Aims and Objectives.....	26
2 Chapter 2: Literature Review	29
2.1 The Origins of Privacy	31
2.1.1 Defining Privacy	32
2.1.2 Alan Westin: Privacy as 'States and Functions'	34
2.1.3 Irwin Altman: Privacy as 'Processes and Goals'	35
2.1.4 Charles Fried: Privacy as Control and Monitoring.....	36
2.1.5 Summary.....	37
2.2 Privacy Regulations in the EU.....	39
2.2.1 From Privacy Protection to Data Protection	39
2.2.2 EU Data Protection: Core Principles	42
2.2.3 Interpreting the EU Directive: Working Party 29	44
2.2.4 Privacy Regulation in the US	45
2.2.5 Managing Transnational Data/Information Flows.....	48
2.2.6 'Safe Harbor' Principle	48
2.2.7 Regulation v's Free Market Approaches	50
2.2.8 Regulatory Environment: Fit for Purpose?	52
2.3 Privacy Policies	53
2.3.1 Machine Readable Privacy Policies	54
2.3.2 Regulatory Changes Proposed	56
2.3.3 Regulation Complexity: The EU Cookie Law	58
2.3.4 Adequate Provisions – Poor Enforcement?	60
2.3.5 Differing Regulatory Approaches – Common Issues.....	61
2.4 Changes to EU Data Protection: Overview.....	62
2.4.1 Transparency	63
2.4.2 Sensitive Data	63
2.4.3 Control.....	64
2.4.4 Privacy Seals	65
2.4.5 US Federal Trade Commission (FTC) - Data Protection Proposals: Overview.....	66
2.4.6 US Department of Commerce: Data Protection Proposals.....	67
2.4.7 The 'White House Report' 2012	68
2.4.8 Summary.....	70
2.5 Iconography / Privacy Seals.....	72
2.5.1 Icons, Labels and Privacy Seals.....	72
2.5.2 Offline Iconography: Energy and Nutrition	74
2.5.3 Online Iconography: Creative Commons Licences.....	76
2.5.4 Online Privacy Seals	77
2.5.5 Privacy Bird	80
2.5.6 Privacy Icons	80
2.5.7 Summary.....	82
2.6 Media Economics Approaches to Privacy	84
2.6.1 Media Markets.....	84
2.6.2 Media Regulation	87
2.6.3 Economic Approaches to Privacy.....	88
2.6.4 Privacy, Advertising and Electronic Commerce.....	94
2.6.5 'Nothing to hide nothing to fear'	96
2.6.6 Information Asymmetry.....	97
2.6.7 Enhancing User Control & The 'Privacy Paradox'	99

2.6.8	Privacy Seal Empirics	102
2.6.9	Privacy Icon Empirics.....	106
2.6.10	Icons and Seals: Summary.....	107
2.6.11	Media Economics: Summary:.....	110
2.7	Chapter Summary and Research Gap:.....	111
3	Chapter 3: Methodology	114
3.1	Research Philosophy	118
3.1.1	Ontology	118
3.1.2	Epistemology.....	120
3.2	Research Methodology	121
3.3	Research Methods	123
3.3.1	Online Survey Method	124
3.3.2	Experiment Method	127
3.4	Sample.....	133
3.4.1	Sample Recruitment	133
3.5	Research Design.....	138
3.5.1	Online Survey Design	138
3.5.2	Survey: Questions.....	139
3.6	Experimental Research Design.....	141
3.6.1	Experiment Design.....	142
3.7	Data Analysis.....	151
3.7.1	Descriptive Statistics.....	151
3.7.2	Inferential Statistics.....	151
3.8	Validation of Data	154
3.8.1	Internal Validity	154
3.8.2	External Validity	155
3.8.3	Validity Tests.....	157
3.8.4	Social Desirability Bias.....	158
3.8.5	MTurk Data Validity.....	158
3.8.6	MTurk: Limitations.....	161
3.8.7	Ethics.....	163
4	Chapter 4: Findings and Analysis	164
4.1	Research Objective 1: Analysis and Discussion	165
4.1.1	RO1: 'Sensitive Data' Survey Findings	166
4.1.2	RO1 Survey: Demographics.....	166
4.2	Research Objective 2: Analysis and Discussion	180
4.2.1	RO2 Sensitive Data: Survey	182
4.2.2	RO2 Sensitive Data: EU Findings	184
4.2.3	RO2 Sensitive Data: US Findings	200
4.2.4	RO2 EU v US Survey: Analysis & Discussion.....	215
4.2.5	RO2: Summary	216
4.3	Research Objective 3.....	218
4.3.1	Addressing Information Asymmetry.....	218
4.3.2	Iconography & New Media.....	219
4.3.3	Privacy Seals: RO3 Research Findings	220
4.3.4	Summary.....	222
4.3.5	RO3: Findings, Analysis & Discussion	224
4.3.6	RO3 - Key Findings: Analysis & Discussion.....	227
5	Chapter 5: Conclusions	229
5.1	Research Objective 1: Conclusion.....	230
5.2	Research Objective 2: Conclusion.....	231
5.3	Research Objective 3: Conclusions	236
5.3.1	Implications for Media Firms.....	237
5.3.2	Implications for Media Regulators	238
5.4	Original Contribution to Knowledge	240
5.5	Concluding Remarks	241
5.6	Limitations:.....	244

5.7	Future Research:.....	244
	References	246
6	Appendices	273
6.1	Appendix 1: RO1 Survey.....	273
6.2	Appendix 2: RO2 Survey.....	274
6.3	Appendix 3: RO3 – EU Experiment	275
6.4	Appendix 4: RO3 – US Experiment	276

List of Figures

Figure 2-1 Academic Definitions of Privacy	Error! Bookmark not defined.
Figure 2-2 FTC Data Protection Principles	Error! Bookmark not defined.
Figure 2-3 Energy Label Example	Error! Bookmark not defined.
Figure 2-4 Creative Commons Licence.....	Error! Bookmark not defined.
Figure 2-5 TRUSTe Privacy Seal	Error! Bookmark not defined.
Figure 2-6 ePrivacy Seal.....	Error! Bookmark not defined.
Figure 2-7 EuroPriSe Privacy Seal	Error! Bookmark not defined.
Figure 2-8 Privacy Bird.....	Error! Bookmark not defined.
Figure 2-9 Disconnect Privacy Icons	Error! Bookmark not defined.
Figure 2-10 Privacy 'Nutrition' Style Label	Error! Bookmark not defined.
Figure 3-1 Trade-off of Different Recruitment Methods	131
Figure 3-2 RO1 Survey: Sensitive Data (EU)	136
Figure 3-3 RO2 Survey 1: Sensitive Data (EU)	136
Figure 3-4 RO2 Survey 2: Sensitive Data (US)	136
Figure 3-5 RO3 Experiment 1 - Privacy Seal (EU).....	137
Figure 3-6 RO3 Experiment 2 - Privacy Seal (US).....	137
Figure 4-1 RO1 - Gender (EU).....	166
Figure 4-2 RO1 – Age (EU)	167
Figure 4-3 RO1 – Education (EU).....	168
Figure 4-4 RO1 – Home Address Details (EU)	169
Figure 4-5 RO1 – Race / Ethnic Origin (EU)	170
Figure 4-6 RO1 – Mobile Phone Number (EU)	171
Figure 4-7 RO1 – Religious Beliefs (EU).....	172
Figure 4-8 RO1 – Email Address (EU)	173
Figure 4-9 RO1 – Health Data (EU).....	174
Figure 4-10 RO1 – Employment Status/History (EU).....	175
Figure 4-11 RO1 – Sexual Orientation (EU).....	176
Figure 4-12 Existing Sensitive Data Categories	182
Figure 4-13 'New' Sensitive Data Categories	183
Figure 4-14 RO2 - Gender (EU)	184
Figure 4-15 RO2 - Age (EU)	185
Figure 4-16 RO2 - Education (EU)	186
Figure 4-17 RO2 – Race / Ethnic Origin (EU)	187
Figure 4-18 RO2 – Political Opinions (EU).....	188
Figure 4-19 RO2 – Religious Beliefs (EU).....	189
Figure 4-20 RO2 – Trade Union Membership (EU).....	190
Figure 4-21 RO2 – Health or Medical History (EU)	191

Figure 4-22 RO2 – Sexual Orientation (EU).....	192
Figure 4-23 RO2 – Physical Location (EU)	193
Figure 4-24 RO2 – Date-of-birth (EU).....	194
Figure 4-25 RO2 – Home Address / Postcode (EU)	195
Figure 4-26 RO2 – Employment (EU).....	196
Figure 4-27 RO2 – Personal Income (EU)	197
Figure 4-28 RO2 – Gender (US).....	200
Figure 4-29 RO2 – Age (US)	201
Figure 4-30 RO2 – Education (US).....	202
Figure 4-31 RO2 – Race / Ethnic Origin (US)	203
Figure 4-32 RO2 – Political Opinions (US).....	204
Figure 4-33 RO2 – Religious Beliefs (US).....	205
Figure 4-34 RO2 – Trade Union Membership (US).....	206
Figure 4-35 RO2 – Health or Medical History (US)	207
Figure 4-36 RO2 – Sexual Orientation (US).....	208
Figure 4-37 RO2 – Physical Location (US)	209
Figure 4-38 RO2 – Date-of-birth (US).....	210
Figure 4-39 RO2 – Home Address (US).....	211
Figure 4-40 RO2 – Employment (US).....	212
Figure 4-41 RO2 – Personal Income (US)	213
Figure 4-42 RO3 – EU Data: Fisher Exact Test Results.....	225
Figure 4-43 Privacy Seal Image Used In EU Experiment	225
Figure 4-44 RO3 – US Data: Fisher Exact Test Results.....	226
Figure 4-45 Privacy Seal Image Used In Us Experiment.....	226
Figure 6-1 RO1 – Example Survey Screenshot	273
Figure 6-2 RO2 – Example Survey Screenshot	274
Figure 6-3 RO3 – EU Experiment Example Screenshot.....	275
Figure 6-4 RO3 – US Experiment Example Screenshot.....	276

List of tables

Table 4-1 RO1 (EU) Responses Ordered By Sensitive Percentages.....	177
Table 4-2 Summary of RO2 Responses (EU).....	198
Table 4-3 RO2 – Responses Ranked (US)	214
Table 4-4 Differences between US and EU Survey Responses.....	215

Acknowledgements

I would like to thank my supervisory team, Dr John Oliver, Dr Kris Erickson and Professor Martin Kretschmer for all of their assistance, guidance and support in writing my thesis. I am also grateful for the financial assistance given by CREATE Glasgow University and the Advances in Media Management research group at Bournemouth University.

Huge thanks to my wife Caz for all her proof reading and feedback and to my sons Louis, Sylvester and James for being so great.

Authors Declaration

This thesis is all original work created solely by myself, Conor O'Kane.

1 Chapter 1: Introduction

This introduction chapter sets out how advances in technology have facilitated the emergence of a global new media industry. Technological innovation has facilitated the delivery of highly personalized digital services media services to consumers of new media (Küng, Picard and Towse 2008). While many of these digital products have proved to be very popular with users, they also pose potential risks for service users. Globally, data protection regulators struggle to keep up with the sheer pace of technological advances in this area. We have even seen the CEO of Apple calling for more regulation to provide reassurances to new media consumers and protect the future of this industry.

In 2018, the EU introduced a comprehensive revision of its data protection provisions in the form of the General Data Protection Directive (GDPR). This thesis conducts experimental research related to a 'new' provision in the GDPR; the establishment of the privacy seal accreditation system.

Leading media economics academic Robert Picard defines media economics as a discipline that addresses,

“how media operators meet the informational and entertainment needs of audiences, advertisers and society” (Picard 1989, p.7).

New media firms technology platforms enable advertisers to efficiently target specific audiences and this has seen firms like Facebook and Google dominate online advertising revenues. Doyle (2013) considers economics as an ideal discipline to view the media sector as,

“most decisions taken by those who run media organizations are, to a greater or lesser extent, influenced by resource and financial issues” (Doyle 2013, p.1),

New media scholars Napoli and Roepnack (2018) state that,

“Concerns about data privacy have been a defining characteristic of the digital media age” (Napoli and Roepnack 2018, p.410).

This thesis examines these and related issues through the lens of media economics.

1.1 New Media Privacy Issues in Context

Apple’s chief executive Tim Cook declared privacy to be one of the most important issues of the 21st century. In a 2018 interview Mr Cook said,

“I’m not a pro-regulation kind of person, I believe in the free market deeply [but] when the free market doesn’t produce a result that’s great for society, you have to ask yourself: What do we need to do? And I think some level of government regulation is important.” (Thornhill 2018).

During his keynote address at the 40th International Conference of Data Protection and Privacy Commissioners (ICDPPC) in November 2018, Cook gives us some insight into why he believes additional regulation of the new media industry is needed. The Apple CEO describes how the trade in digital data has exploded into what he calls a ‘data industrial complex’. He somewhat startlingly sets out how,

“Our own information — from the everyday to the deeply personal — is being weaponized against us with military efficiency.....These scraps of data, each one harmless enough on its own, are carefully assembled, synthesized, traded and sold. Taken to the extreme this process creates an enduring digital profile and lets companies know you better than you may know yourself. Your profile is a bunch of algorithms that serve up increasingly extreme content, pounding our harmless preferences into harm.” He goes on to what that, “We shouldn’t sugarcoat the consequences. This is surveillance,” (Lomas 2018).

It is not entirely surprising that new media firms go to such lengths to collect and synthesize data. In 2017, the Economist editorial declared data as the new oil.

“A new commodity spawns a lucrative, fast-growing industry, prompting antitrust regulators to step in to restrain those who control its flow. A century ago, the resource in question was oil. Now similar concerns are being raised by the giants that deal in data, the oil of the digital era. These titans—Alphabet (Google’s parent company), Amazon, Apple, Facebook and Microsoft—look unstoppable.” (Economist 2017).

The editorial also calls for greater transparency by forcing these new media service providers into disclosing what information they hold on their customers.

Edward Snowden is the former US National Security Agency (NSA) contractor and latter whistleblower, who leaked top-secret NSA documents detailing phone and Internet surveillance practices by US security agencies. In a 2016 interview, Snowden stated,

“[In the past,] your beliefs, your future, your hopes, your dreams belonged to you,”.....“Increasingly, these things belong to companies, and these companies can share them however they want, without a lot of oversight” (Ha 2016).

While Snowden was not advocating that companies should completely stop collecting user data, he was stating his belief that a central problem for the future was that individuals themselves needed to be in control of their own data.

The World Economic Forum (WEF) believes the collection and storing of large amounts of individuals personal information poses significant macroeconomic challenges for the world. In its ‘Global Risks Report 2017’, the WEF identifies a number of ‘global risks’ to the world economy. It lists them in 5 categories covering economic, environmental, geopolitical, societal and technological risks. The report defines ‘global risks’ as,

“an uncertain event or condition that, if it occurs, can cause significant negative impact for several countries or industries within the next 10 years” (WEF 2017, p.62).

The report further identifies the ‘Top 5 Global Risks in Terms of Likelihood’ (i.e. most likely to occur in the next 10 years) and identifies a, ‘*Massive incident of data fraud/theft*’ as a top 5 ‘likelihood’ risks (WEF 2017 pg4). It is worth noting this was the first time the WEF listed a ‘massive incident of data theft/fraud’ as a top 5 likely events. As a technological risk, it is also worth noting a ‘massive incident of data fraud/theft’ is considered more likely event than ‘large scale cyber-attacks’. So on a macro level, data theft/fraud is considered to have the potential to have a significant negative impact on the global economy and is listed alongside risks including, ‘extreme weather events’, ‘large-scale involuntary migration’, ‘major national disasters’ and ‘large scale terrorist attacks’. While network security is not the focus of the research being described here, the WEF report does highlight the need for individuals to be more vigilant in terms of when and who they share their personal information with.

No internationally agreed consensus exists on common standards for how media corporations or governments protect personal information. Academic scholars in new media warn that that global nature of how new media firms operate means that,

“collected data also might easily be transported across national borders to be archived and processed in a location where the individual’s expectations of privacy and data security, based on his or her home country laws, may not be fulfilled” (Alabarran, Mierzejewsha and Jung 2018, p.192).

While one obvious solution to this problem is would be for counties to prohibit (in law) the transfer of citizens’ data, this has the potential to limit economic growth in countries that introduce such rules. In the EU, the Privacy Shield principle (examined in more detail in later chapters) puts in place the legal requirement that firms who transfer personal information of EU citizens to another jurisdiction (for processing or storage purposes etc) must adhere to EU data protection provisions.

The 2016 Brexit referendum in the UK and the 2016 presidential elections in the US saw election campaigns successfully target potential voter groups highly personalized messages,

“We have no way of knowing how our personal data is being mined and used to influence us. We don’t realise that the Facebook page we are looking at, the Google page, the ads that we are seeing, the search results we are using, are all being personalised to us. We don’t see it because we have nothing to compare it to. And it is not being monitored or recorded. It is not being regulated. We are inside a machine and we simply have no way of seeing the controls. Most of the time, we don’t even realise that there are controls.” (Cadwalladr 2016).

In the UK, the Information Commissioners Office (ICO) fined Facebook the maximum permitted £500,000 for,

“allowing third party developers to access user information without sufficient consent” (Hern and Waterson 2018).

Facebook were found to have failed to keep personal information about their users safe and although in total 87 millions users data was ‘harvested’ the ICO fine relates only to the 1 million UK users affected.

“The fine is for two breaches of the Data Protection Act. The Information Commissioner’s Office (ICO) concluded that Facebook failed to safeguard its users’ information and that it failed to be transparent about how that data was harvested by others.” (Hern and Pegg 2018).

So how did we get here?

1.2 The Rapid Emergence of ‘New Media Industry’

The Oxford English dictionary defines the term ‘new media’ as,

“means of mass communication using digital technologies such as the Internet” (Oxford Dictionaries 2018).

Traditionally, media firms were defined as enterprises, “involved in the production and distribution of content intended for a mass audience” (Mierzejewska and Shaver 2014, p48). However, media scholars suggest that research should focus on

firms that produce content as well as firms who operate platforms that distribute content, including user based content. This means that,

“companies such as YouTube, Netflix, Facebook, Google and Apple become as important to media management research as traditional mass-media firms” (Rohn 2018, p.430).

Technological growth is the key driver of a rapid expansion in the global ‘new media’ industry in recent years (Horst and Murschetz 2019; Küng 2017, Oliver 2018b; Oliver and Picard 2020). As such,

“Economic success in the media industry is naturally dependent on the ability to adjust and capitalize on technological advances” (Doyle 2013, p.26).

New media firms including Alphabet Inc. (Google parent company), Facebook, Microsoft and Apple (as well as a host of others) have grown to become global brands in relatively short periods of time. The real and virtual products and services they offer have been popular with both users and investors alike.

Alphabet was founded in 1994 and figures for 2017 show annual revenues in excess of \$110 billion (Alphabet Investor Relations 2018). Their ‘search’ product dominates its rivals where it accounts for approx. 92% of online searches worldwide (StatCounter 2019). This creates the platform for its AdWords product, allowing advertisers to target users based on their online activities and generates a staggering 90% of its total annual revenue.

Facebook founded 2004, in Dec 2018 Facebook had 1.52 daily active users and 2.32 billion active monthly users (Facebook 2019). At its Initial Public Offering (IPO) in 2012 it was valued at \$104 billion (WSJ 2012). Despite losing £123 billion in value as a result of a major data breach, July 2018 saw the company still valued at a staggering \$630 billion. Facebook does not charge a membership fee and its primary income driver is advertising revenues. Figures for year ending Dec 2017 saw its revenues earned from advertising hit \$39.9bn, a rise of 48% from the previous year and with 98% of its revenue coming from advertising (Statista 2018).

“The ascent of the so-called Web 2.0 (blogs, social media, online social networks) has rendered individuals no longer mere consumers of information, but public producers of often highly personal data.” (Acquisti, Taylor and Wagman 2016, p.444).

Instead of users’ data being stored locally on devices, digital information (emails, documents, music, chats, photos etc) is stored in a digital ‘cloud’ in the form of vast data centers. The data is then accessible through the range of networked devices including PCs, mobile phones, iPods etc. Some commentators suggest,

“the potential power of these cloud providers raises all sorts of fears, not least for privacy, security and reliability” (Leadbeater 2010).

New media firms have also attracted criticism from privacy campaigners who claim that elements of the way some products and services operate constitutes an invasion of privacy; and that the advertising incomes they attract are as a result of disclosure of users’ personal information.

Leading privacy-economics academic Alesandro Acquisti describes the impact new media services and the challenges this brings as follows:

“New search engines, social networks, ecommerce websites, web browsers, and individualized controls for privacy-conscious consumers have emerged. Concurrently, social media services have facilitated a culture of disclosure: a disclosure of one’s activities, location, emotions, work history, and political opinions. While, overall, these technologies seemingly leave privacy choices in the hands of consumers, many (if not most) consumers, in practice, lack the awareness and technical sophistication required to protect and regulate the multiple dimensions of their personal information. Privacy-invasive technological services have become integral to every-day communications, job searches, and general consumption. At the same time, privacy-protecting services require additional levels of user effort and know-how, which limits their efficacy, especially within some of the most vulnerable segments of the population” (Acquisti et al. 2016, p.484).

What evidence is there to support the argument that social media services facilitate a 'culture of disclosure' or that consumers 'lack the awareness and technical sophistication' to regulate their personal information when using these new media services?

1.3 New Media: A Culture of Disclosure?

Indeed, there is evidence that new media firms promote a so-called culture of disclosure. Facebook is no stranger to such controversy in relation to privacy. As far back as 2007, its ad serving Beacon service hit the headlines when it was found to have failed to inform users on how it shared their details with other services. The controversy resulted in a class action lawsuit and a \$19m fine. In 2017 Facebook was fined €110m for, "misleading the European Commission during its 2014 takeover of WhatsApp" (Murgia 2017). Facebook had explicitly told EU regulators that it could not automatically link Whatsapp customer accounts with its own platform and yet did this shortly after the acquisition and was therefore found to be in breach of EU merger rules. Facebook was further separately fined by French and Italian regulators over the way it tracked users web browsing, making this information available to advertisers, in breach of EU data protection laws.

Academic research looking specifically at how Facebook operates has suggested that although the company has refrained from explicitly selling users' information directly to advertisers,

"they have created systems that enable advertisers to run highly targeted social advertising campaigns" (Korolova 2010, p.474).

The rapid growth in revenue earnings would suggest that this indeed is the case. Facebook founder Mark Zuckerberg challenges privacy campaigners arguing that privacy is no longer a 'social norm' and that,

“People have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people” (Barnett 2010).

However, Mr Zuckerberg’s comments may be more grounded in commercial imperatives rather than representing a fundamental paradigm shift in how millions of people around the world view their personal privacy.

The following provides a useful example of how Facebook does not directly sell users personal details although it can perhaps been seen as facilitating a ‘culture of disclosure’. Voncom is just one of a multitude of firms that provide interactive content on the social network including quizzes, top 10 lists etc. In 2015, Voncom launched an interactive service that allowed Facebook users to create a ‘info-graphic’ or ‘world cloud’ based on the words you had used most often on the social network. When a user clicked on the link to use the service, they were shown a message to say that the app would need to access their Facebook data in order to create the ‘word cloud’ image. It was a popular app used by and estimated 17+ million users (Wakefield 2015).

As it transpired, the personal data the ‘info-graph’ application accessed included the users,

“name, profile picture, age, sex, birthday, entire friend list, everything you have posted on your timeline, all of your photos, home town, education history and everything you have ever liked” (IBID).

There is no evidence or accusation that Facebook or Voncom breached data protections rules as the terms and conditions of the app included accessing their Facebook data. Voncom says it only accesses users personal data to ‘improve the quality of its services’ and that it does not sell users personal information to third parties, even though this is permitted in its terms and conditions. However, this example highlights a potential lack of awareness individuals have when using ‘privacy invasive’ applications. Would 17m people have used the app had they known how much personal information about themselves they were disclosing?

More recently, a case taken by Austrian student Max Schrems resulted in the European Court of Justice (ECJ) ruling that the 'Safe Harbour' principle, a legal agreement that allowed companies collecting individuals personal information (including the likes of Facebook, Alphabet and Microsoft etc) to transfer EU citizens' data to the US for processing, was no longer valid. It ordered that the transfer of data to the US relating to Facebook's European subscribers be,

“suspended on the ground that that country does not afford an adequate level of protection of personal data” (Court of Justice of European Union 2015, p.3).

While this ruling from the ECJ was specific to Facebook, it has wider implications for all media firms reliant on the Safe Harbour principle. The US Federal Trade Commission commissioner Julie Brill described the impact of the ECJ ruling as, “measuring 7.8 on the Richter scale” (O’Kane 2015). The Safe Harbor Principle has since been replaced with a similar arrangement known as Privacy Shield, which is subject to ongoing legal challenges.

Googles parent company Alphabet Inc. has also attracted controversy in relation to privacy. While collecting data for their 'Street View' service, it emerged they had been scanning and collecting information from Wi-Fi networks within range of their vehicles. Commentators claimed that Alphabet's actions amounted to a, “violation of customers trust” (Metz and Goodin, 2010). They attracted further criticism from privacy campaigners in relation to default settings on its Buzz social media service. Alphabets then CEO Eric Schmidt was questioned on these issues and the interviewer asked if users should be prepared to share information with them as, “a trusted friend”. Schmidt replied saying,

“If you have something that that you don't want anyone to know, maybe you shouldn't be doing it in the first place” (Gelles et al 2009, p.11).

Issues have also emerged in relation to relatively complex arrangements between new media companies over ownership of the information they hold on

users. In May 2013, UK media telecoms operator EE attracted the attention of privacy campaigners when it emerged that a separate company, Ipsos Mori, had an exclusive deal to sell on EE customers data. The London Metropolitan Police confirmed they had met with Ipsos Mori to discuss how data could potentially be used in crime prevention. Information on approximately 27m customers was available, although EE confirmed that they were fully in compliance with the UK data protection regulations. EE stated the data available was anonymised and could not be used to identify individuals. Personal data that was available on customers included gender, age, postcodes, websites visited and locations of customers when making calls (Lee, 2013).

Indeed, a recent investigation by the UK consumer group Which? found that,

“sensitive personal and financial data is being traded on a huge scale, with some companies apparently willing to sell to anyone who comes calling” (McLaughlan 2017).

Which?’s undercover investigators were offered 500,000 items of personal information, including phone numbers and addresses on households with an income above £40,000 for just four pence each. Perhaps most interesting was that none of the companies contacted performed any due diligence on the fake pensions company Which? was using to make the data requests through. The UK Information Commissioners Office is investigating to see if data protection rules were broken.

The privacy issues discussed above are all examples where new media firms have found themselves embroiled in controversy in relation to their data protection practices. So what regulatory provisions exist to protect media consumers against threats to their privacy? The following section will give a brief overview of the key relevant privacy regulations in the EU and US.

1.4 Regulating Privacy

Both the European Union (EU) and United States of America (US) have adopted contrasting approaches in terms of how they regulate data protection in their respective jurisdictions. They are also the largest and wealthiest economies in the world collectively accounting for almost 50% of global GDP.

In recent years, the relevant EU and US data protection authorities have outlined how rapid technological advances pose challenges to existing data protection regulations. In 2012, the European Commission published a communication to the European Parliament declaring,

“rapid technological developments and globalisation have profoundly changed the world around us, and brought new challenges for the protection of personal data” (European Commission 2012, p.2).

The challenge for data protection regulators is to strike a balance between protecting individuals right to personal privacy and providing a business/commercial environment that promotes economic growth.

A number of international agreements include privacy as a fundamental right that all humans are entitled to. These include Article 12 of the Universal Declaration of Human Rights (UDHR) (United Nations, 1948) and Article 8 of the European Convention on Human Rights (ECHR) (Council of Europe, 1950). Both treaties state that all individuals have a right to privacy protected in law. Informed by these treaties, European Union and individual members’ data protection regulations have developed to protect individuals’ privacy rights. In the European Union (EU), a group called the Working Party 29 (WP29) provide expert opinion to the European Commission (EC) on questions/issues that arise in relation to data protection.

“An examination of past, present and future data protection legislation reveals that the protection of personal data evolved from the right to private life as provided for in Art. 8 of the ECHR of 1950” (Working Party 29 2014, p.3).

WP29 believes that rapid advances in technology has created the requirement for further protection of individuals right to control their personal data, and have stated,

“With the increase of new technologies and surveillance possibilities, both in the public and in the private sector, became apparent that there needed to be further protection for individuals from third parties (particularly the State) in addition to ‘defensive’ rights recognised under Art. 8 of the ECHR by ensuring that he individual had the right to control his/her own personal data.” (Working Party 29 2014, p.3).

There are similar data protection concerns in the US where the two bodies responsible for data protection, the Federal Trade Commission (FTC) and the Department of Commerce (USDoC) published major reports in 2010 pointing to serious failings existing data protection provisions.

“As the ability of corporations and government to collect, archive and process individual data has expanded exponentially, no consensus has emerged on common standards of protection for personal information” (Jayakar 2018 p.191-192).

This is not entirely surprising as regulators face a complex task when developing data protection regulations, not least because,

“privacy sensitivities and attitudes are subjective and idiosyncratic, because what constitutes sensitive information differs across individuals” (Acquisti et al. 2016, p.446).

Data protection regulators have for a number of years expressed concerns in relation to the protection of personal information. The EU describes how advances in technology have created the requirement for further protection of individuals’ right to control their personal data.

“With the increase of new technologies and surveillance possibilities, both in the public and in the private sector, became apparent that there needed to be further protection for individuals from third parties.....by ensuring that the individual had the right to control his/her own personal data.” (Working Party 29, 2014 p.3).

In response to the 2012 Europeans Commission calls for a comprehensive data protection package, the Council of Europe on the 15th December 2015 published a, ‘*Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*’ (Council of European Union, 2015). This General Data Protection Regulation (GDPR) replaces the EU Data Protection Directive 94/46/EC (Data Protection Directive). The GDPR was adopted in April 2016 and became effective (and directly applicable) in all EU states from May 2018. The GDPR updates existing provisions as well as introducing new concepts that did not exist under the Data Protection Directive.

1.5 Research Gap

Issues around data protection and privacy have received limited attention among academics researching in the media economics and media management discipline. This is evidenced by the fact that a leading text Handbook of Media Economics and Management 2018 text, a publication that,

“effectively constitutes a kind of “best of” media management and economics that very well summarizes the main issues and trends of the field” (Jung et al, 2018, p.xiv)

devotes just over half a page of text (out of 450 pages) to the topic of privacy in the context of data.

In the limited section that does address privacy, it is noted that,

“it is important to acknowledge the privacy implications of a media environment in which managerial decision-making is increasingly dependent on the gathering and analysis of large quantities of personal data” (Napoli and Roepnack 2018, p.417).

However, somewhat surprisingly, the chapter also states that,

“The days of widespread consumer unease with the ever-increasing forms of data gathering that the media sector engages in seems to have passed” (IBID).

The idea that media consumers are no longer uneasy with how media firms collect and manage their personal data is certainly out of step with how EU and US regulators view this issue. It also suggests that a research gap exists within the media economics and media management fields and this thesis will address some key elements that new media operators need to address in relation to relevant regulatory provisions.

The GDPR sets out all of the provisions for,

“the protection of natural persons with regards to the processing of personal data and on the free movement of such data” (GDPR 2016, p.1).

Recital 100 proposes,

“In order to enhance transparency and compliance with this Regulation, the establishment of certification mechanisms, data protection seals and marks should be encouraged, allowing data subjects to quickly assess the level of data protection of relevant products and services”, (GDPR 2016, p.19).

It is this proposed use of privacy seals as a privacy enhancer that is the focus of the research conducted in this thesis. There is relatively little empirical evidence that examines how privacy seals (and/or related iconography) impacts on personal information disclosure. Our current understanding of privacy seals in academic literature (Tsai, Egelman, Cranor, and Acquisti, 2011; Brandimarte, Acquisti, and Loewenstein, 2012; Carolan and Castillo-Mayen 2014) suggests

that privacy seals/icons do not necessarily operate in a way that regulators might assume. Findings show inconsistent and at time somewhat counter-intuitive behavior from data subjects when these 'transparency enhancing' features are displayed. Greater clarity in relation to the role privacy seals can play in promoting information transparency has implications for data subjects (as service users) and data controllers (as service providers) and policy makers (as industry regulators). With many new media firms so dependent on advertising revenues, a third party-verified accreditation scheme that confirms compliance the GDPR has the potential to gain the confidence of data subjects and in turn may lead to increased revenues. Alternatively, if privacy seals have a negative effect on personal information disclosure, then new media firms may choose not to participate in privacy seal accreditation schemes resulting in the improved transparency desired by regulators not being achieved.

The EE data ownership example touches at the core of one of the key areas of concern i.e. a need to make it clear to individuals how and to whom their personal information is being shared. Regulators generally refer to this as clarity around what personal information is being collect and how it is used share as 'information transparency'. The following chapters will examine literature from a range of disciplines including media economics, economics and privacy theory exploring the key characteristics and complexities in relation to personal information disclosure. After examining key areas where data protection regulators say existing regulatory provisions need enhancing, a number of online experiments will be set out to examine the effects these privacy enhancing features may have on users' information disclosure.

1.6 Aims and Objectives

It is clear from the growth rates, valuations and revenue streams of many new media firms that the collection and harvesting of users personal data including their click-through, photos, location data, networks of friends etc have substantial economic value. In an attempt to keep up with technological developments the EU has introduced a comprehensive new regulation in the form of the GDPR. However, this is a new regulation and as such the

enforcement mechanism (for breaches of the regulation) remain untested to date. For data subjects, there is a risk to their privacy through the collection and sharing practices of new media firms. For data controllers, apart from potential regulatory fines, breaches of data subjects privacy may result in a loss of confidence in these firms resulting in a negative impact on their business models/revenues. The WEF even consider a large theft of personal data both a likely and substantial risk to the global economy.

The GDPR represents an attempt by regulators to bring data protection provisions up to date with technological innovations. Improving transparency and increasing the control individuals have over their personal data are key objectives of this new regulation. The GDPR explicitly calls for the establishment of a privacy seal accreditation system where the display of an accredited seal may be used to demonstrate compliance with the new regulation. However, there is little empirical evidence in relation to whether privacy seals improve transparency in the way regulators desire. The research undertaken here uses online behavioral experiments in an attempt to help fill this research gap.

The overarching research question this thesis addresses is the following:

What impact(s) do Privacy Seals have on sensitive personal information disclosure in online environments?

In order to answer this research question, the following research objectives are addressed:

RO1 – To examine if respondents can identify the types of data categorised as ‘sensitive data’ status under current EU data protection regulations.

RO2 – To examine what categories of personal data respondents consider 'sensitive personal data'.

RO3 – Conduct experiments to test if use of 'privacy seals' effect data subjects personal information disclosure.

2 Chapter 2: Literature Review

For many new media firms, a very significant percentage of total revenues come from advertising income. Technology has revolutionised the new media-advertising sector and behavioural advertising is a key driver of this. Online advertisers use personal data available on new media platforms to deliver personalised content and products to consumers (Küng, Picard and Towse, 2008). We also know that,

“Advertisers are critical to the success of commercial media because they provide the primary revenue stream that keeps most of them viable” (Picard, 2002, p.122).

The metrics that digital advertising offers makes it an attractive area to spend, especially when compared to the rather inexact science of traditional offline advertising methods.

All media products/services must adhere to the regulatory requirements for the jurisdictions relevant to where it operates (Picard 2002). Therefore, any change to EU data protection regulations has the potential to impact the operations and revenues of any new media firms who operate within the EU jurisdiction.

The GDPR represents a comprehensive overhaul of data protection regulations and as such impacts all new media firms that operate services with the lucrative markets that is the EU (Rossi 2017). Recital 100 of the GDPR is the primary concern of this thesis. To recap, it calls for the establishing of Privacy Seal accreditation scheme.

“In order to enhance transparency and compliance with this Regulation, the establishment of certification mechanisms, data protection seals and marks should be encouraged, allowing data subjects to quickly assess the level of data protection of relevant products and services”, (GDPR 2016, p.19).

This chapter examines relevant existing theoretical and empirical literature across a range of disciplines including media economics, economics, law, and

political science. It is organised as follows. While the findings of this research are of most relevance to the media economics discipline, the literature review does not begin with media economics. We begin by exploring academic definitions of privacy, establishing core themes among leading academics. We then examine the founding principles behind the origins of data protection regulations in the EU and US. While the regulatory approaches adopted are clearly different, the issues and challenges these services present in relation to online privacy share many common concerns for their respective regulators. We then examine literature from media economics, including the key role of regulation. We then trace the origins and development of economics literature as it pertains to privacy. Finally, key empirical literature across a range of disciplines that explores the impact privacy seals and similar related iconography has on personal information disclosure is discussed. A clear research gap is identified.

2.1 The Origins of Privacy

While the Bible has numerous references to privacy, early Hebrew, classical Greek and ancient Chinese cultures saw more substantive protection of individual notions of privacy - mostly focused on the right to solitude (Hixson 1987). The law of privacy can be traced back to 1361 when the English Justices of the Peace Act established 'peeping toms' and 'eavesdropping' as offences (Davies 2001). In 1765, a ruling by British Lord Camden struck down a warrant to enter a house to seize papers saying there was no law in the country to justify this. He added that if such a law did exist, "it would destroy all the comforts in society" (Entick v. Carrington, 1558-1774 All E.R. Rep. 45).

The world's first data protection act dates from Sweden in 1776. The '*Access to Public Records Act*' restricted the use of government held information to 'legitimate' purposes only. In 1792 the French '*Declaration of the Rights of Man and Citizen*' saw private property established as inviolable and sacred. Then in 1858, France prohibited the publication of private facts on individuals and established fines for violators (Henderson and Snyder 1999).

Research conducted by Irwin Altman in the 1970's examined a range of cultures (including the cultures of apparently minimal privacy like the Mehinacu tribe in Brazil and the Pygmies of Zaire) and found privacy to be culturally universal phenomena. Altman argues that privacy is fundamental to both individual and wider cultural survival (Altman 1977).

Privacy is a complex and multi-faceted term. Davies (2001) categorises it into four distinct facets: a) *Information Privacy*: Establishing rules of governing the collection of personal data and how this data is managed and handled. b) *Privacy of Communications*: Covers the security of telephones, emails and other forms of communication. c) *Territorial Privacy*: Setting the limits/boundaries on intrusion into domestic, workplace or other public environments. d) *Bodily Privacy*: Concerns the protection of the physical being against potential intrusions such as drug testing and cavity searches. This thesis focuses on Information Privacy, defined by Clarke as an,

“Individuals claim that data about themselves should not be automatically available to other individuals and organisations, and that, even where data is possessed by another party, the individual must be able to exercise a substantial degree of control over that data and its use.” (Clarke 2013).

2.1.1 Defining Privacy

To gain a better understanding of privacy, we must define it. However, there is no universally accepted definition of privacy. Therefore, we examine definitions from a range of disciplines including key contributions from political scientist Alan Westin, social scientist Irwin Altman and legal academic Charles Fried. These theorists are relevant because their contributions have stood the time, featuring prominently in major privacy academic publications over the last 40 years (Margulis 2003).

In what is widely regarded as a seminal article in relation to privacy, the 1890 Harvard Law Journal published a paper entitled ‘*The Right to Privacy*’. It quoted Judge Thomas Cooley’s claim for an individual’s “right to be let alone”. Interestingly, despite writing over 125 years ago, authors Warren and Brandeis specifically make reference to the threat to privacy that technological development presents. They described how,

“Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good that prediction that ‘what is whispered in the closet shall be proclaimed from the house-tops” (Warren and Brandeis 1890, p.195).

They argue there is ‘no doubt’ that protection from such invasions of privacy is a ‘necessity’. They also describe how the press and other branches of commerce, through the invasion of individuals’ privacy, act in a manner where it is in fact the supply of such information in itself creates the demand. Lawyer Charles Fried later provides a similarly concise definition of privacy, “control over knowledge about oneself” (Fried 1968, p.483).

Political scientist Alan Westin defines privacy as,

“the ability of the individual to control the terms under which personal information is acquired and used” (Westin 1967, p.7).

In 2003, Westin ‘expands’ his definition to reflect the new media or ‘internet era’, defining privacy as,

“the claim of an individual to determine what information about himself or herself should be known to others...This, also, involves when such information will be obtained and what uses will be made of it by others” (Westin 2003, p.7).

Social scientist Irwin Altman defines privacy simply as, “the selective control of access to the self.” (Altman 1975, p.18). Altman further argues,

“the ability to regulate interaction is necessary for individual and cultural survival” (Altman 1977, p.82).

The table below summarises the key definitions of privacy from authors discussed above. Control clearly emerges as a common theme across disciplines.

Figure 2-1 Academic Definitions of Privacy

Year	Definition	Reference
1890	"right to be left alone".	(Warren and Brandeis, 1890, p.195)
1967	"the ability of the individual to control the terms under which personal information is acquired and used"	(Westin 1967, p.7)
1968	" control over knowledge about oneself".	(Fried 1968, p.483)
1975	"the selective control of access to the self."	(Altman 1975, p.18)
2003	"...this, also, involves when such information will be obtained and what uses will be made of it by others"	(Westin 2003, p.3)

While on one hand these definitions of privacy are seemingly different, they are related, as they pertain to the boundaries between the individual self and the other, between private, and/or public (Altman 1975). We will examine these approaches to privacy in more detail.

2.1.2 Alan Westin: Privacy as 'States and Functions'

Westin (1967) describes privacy as a set of changing states and functions. The first state of individual privacy is *solitude*, where one is completely free from the observation of others. The second is *intimacy*, where small groups separate themselves from others in order to be alone, perhaps like a husband and wife might do. *Anonymity* occurs where a person is in a public place but expects not to be recognised. Going to the cinema or attending a concert alone might be an example of this. Finally, *reserve* is a state where we limit disclosure to others,

an example being where in the company of others we 'tune out' persons or communications at various times. These various states of privacy effectively act as the "hows" of privacy (Margulis 2003). Within these states an individual's needs are constantly changing. A person may want to be completely alone while soon afterwards want to be in the presence of a partner or close friend, or discussing personal issues with a complete stranger (Westin, 2003).

He also describes four functions of privacy. *Personal autonomy* relates to the desire not to be manipulated or exposed by others. *Emotional release* is where people can relax and take 'downtime' from social restrictions or customs. *Self-evaluation* deals with extracting meaning from personal experiences and planning future actions. Finally, *limited and protected communication* sets interpersonal boundaries and allows for sharing of selected personal information with others who we trust (Westin 1967).

2.1.3 Irwin Altman: Privacy as 'Processes and Goals'

Altman (1977) sees privacy as a set of processes and goals. Firstly, as a *Dynamic Dialectic Process*; he describes privacy as a boundary control process where we change how open or closed we are in response to internal and external conditions. These change over time and are depending on the specific circumstances an individual finds themselves in. Secondly, as *An Optimisation Process*; this describes privacy as not being monotonic i.e. that more privacy is not necessarily better. He proposes the idea of an optimal level of privacy where the desired level of privacy is equal to actual privacy achieved. He also acknowledges possibilities of too little privacy where desired level is greater than actual levels (i.e. social isolation) and perhaps too much privacy where desired levels are less than actual levels achieved (i.e. overcrowding). Thirdly, as a *Multimechanism Process*; here Altman describes,

"a self-other boundary control process, privacy is viewed as involving a network of behavioural mechanisms that people use to achieve desired levels of social interaction" (Altman 1977, p.67).

The mechanisms here include verbal and non-verbal behaviours and he argues that privacy regulation encompasses much more than management of physical environment in relation to social interaction (IBID).

Altman proposes that privacy is concerned with three important goals. The first of these involves managing the relationship with the social world. Secondly, managing the interface between oneself and the social world, and thirdly issues relating to self-definition and self-identity (Altman 1977).

2.1.4 Charles Fried: Privacy as Control and Monitoring

Surveillance and identity theft are common themes that arise when discussing privacy. Here Orwell's seminal '1984' novel depicting the potential for invasive and subversive monitoring of our personal information by Big Brother perhaps comes to mind (Palen & Dourish 2003). American law academic Charles Fried, while writing in the Yale Law Journal in 1968 described the existence of electronic devices,

“so small as to be entirely unobtrusive: other persons cannot tell that a subject is “wired” and even the subject himself – if he could forget the initial installation – need be no more aware of the device than of a small bandage” (Fried 1968, p.475).

Fried argues that it is not just the amount of information privacy is concerned with, but it is also the quality of that data. He cites the example that we might be happy for a casual acquaintance to know that we are sick, but it may constitute a violation of one's privacy for them to know the exact nature of the illness.

For Fried, love, friendship and trust are all identified as important concepts in privacy. Trust is regarded as more functional; that is, we can trust persons whom we neither love nor even like, whereas both friendship and love imply a certain level of trust. Privacy also has a key role to play in protecting our liberty,

“if we thought that our every word and deed were public, fear of disapproval or more tangible retaliation might keep us from doing or saying things which we would do or say if we could be sure of keeping them to ourselves or within a circle of those who we know approve or tolerate our tastes” (Fried 1968, p483).

For Fried, electronic monitoring represents, “an intolerable violation of privacy”, as it dramatically reduces (or even eliminates) our ability to control information about ourselves (Fried 1968, p490). His words resonate with the type of harmful surveillance that Tim Cook warned about in his ICDPPC keynote address in November 2018.

2.1.5 Summary

As we have seen, privacy is a wide-ranging and complex topic. Both emotions and/or social setting impact our levels of desired privacy. In academic definitions of privacy, ‘control’ emerges as a core theme. Westin’s (2003) expanded privacy definition takes into account the impact technology has on *how and when* our personal information is accessed and shared. This definition reflects the challenges to personal privacy posed by the use of networked technologies, a key enabler of the new media industry. The actions of some global new media firms examined in the introduction can be viewed as breaches of these academic definitions of privacy, where individuals lose control over their personal data, thus impacting their privacy.

The reach of ‘new media’ firms like Facebook, Apple and Google makes privacy a concern with respect to their operations across the globe. Revelations in relation to widespread surveillance carried out by US and UK based national security agencies led to comments from Tim Berners-Lee (the widely accredited ‘inventor’ of the Internet) that echo Altman’s comments as Berners-Lee believes that increased surveillance poses a threat to the very future of democracy (Arthur 2013).

Research published in the Harvard Law & Policy Review,

“empirically demonstrate that advertisers are making it impossible to avoid online tracking” (Hoofnagle, Soltani, Good and Wambach 2012, p.273).

This research shows how tracking technologies have moved on from the standard text file ‘cookies’ (that can be deleted by users) to ‘flash cookies’ that can track individuals irrespective of whether they wish to be tracked or not. Flash cookies and similar tracking techniques take control away from individuals and clearly represent a serious challenge to the core components of the academic definitions of privacy. These technological developments led EU and US regulators to undertake major reviews of how existing data protection regulations in light of new media developments.

The discussion in this thesis will address Information Privacy and how personal information disclosure impacts an individual’s privacy. Based on this approach, the most relevant definition of privacy comes from Alan Westin,

“the claim of an individual to determine what information about himself or herself should be known to others...This, also, involves when such information will be obtained and what uses will be made of it by others” (Westin 2003, p3).

This definition captures the two key objectives of the landmark EU 95/46/EC Data Protection Directive (later incorporated into the GDPR). Firstly, respecting the individuals right to privacy (i.e. the claim of an individual to determine what information about himself or herself should be known to others). Secondly, it addresses how this information is obtained and used by third-parties as well as how media businesses manage this ‘free flow’ of information i.e.

‘when such information will be obtained and what uses will be made of it by others’. (Directive 95/46/EC, 1995)

We will begin by examining the origins and key principles of EU data protection directives.

2.2 Privacy Regulations in the EU

“The roles that governments play in economic affairs are necessary to create and perpetuate markets” (Picard 2002, p.69).

Although notions of privacy have existed for many centuries, the experiences of WWII are the driving force behind European data protection laws (Loring 2002). Elaborate population registers in Holland, compiled primarily for public policy and state planning purposes were used by the invading Germans to identify individuals for deportation, many to concentration camps. Here, while the information itself was not harmful, it serves as an example of where personal information freely supplied today for one reason, may be looked upon and treated in a very different way in the future (Lloyd 2011).

The immediate post-war period in Europe saw the creation by treaty in 1949 of the Council of Europe. This led to the establishing of the European Convention on Human Rights that came into force in 1953. Article 8 specifically related to privacy and states that:

1. Everyone has the right to respect his private and family life, his home, and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as in accordance with the law and it necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

(Council of Europe 1950)

2.2.1 From Privacy Protection to Data Protection

While the Human Rights convention established a right to personal privacy, the 1960's saw Europe progress from an industrial economy to an information economy. This heightened concerns about privacy in relation to personal information/data (Loring 2002).

In parallel to individual countries enacting domestic legislation, 1973 saw The Council Of Europe's intergovernmental body (the Committee of Ministers) pass Resolution 22 covering the 'ground rules' for data protection in the private sector. A year later Resolution 29 was passed, focussing on the public sector (Reed 2011). Resolution 22 and 29 urged individual member nations to adopt data protection legislation.

Indeed, the 1970's saw the first data privacy legislation become law. The German state of Hesse enacted the first data protection law in 1970. Sweden introduced its own data protection legislation in 1973 and was followed by Germany in 1977 and France in 1978. However, divergent privacy standards began to appear between member states, raising concerns that differences in national legislation may pose significant barriers to the transmission of data between European nations (Murray 1998). Divergence was also seen as having a negative impact on the key strategic goal of creating a single market.

In an attempt to create a unified data protection environment and promote greater economic ties among member states, two important sets of international guidelines emerged. These were:

1. ***The Convention for the Protection of Individuals with Regards to the Automatic Processing of Personal Data***: This Council of Europe established convention came into force in 1985. It had five countries as signatories and established a number of basic provisions in relation to data protection. The aim was to provide a minimum degree of harmonisation between signatories in order to eliminate restrictions on data flows between countries for reasons related to data 'privacy' protection. The major weakness and criticism of the Convention is its lack of enforceability, as no enforcement agency was created leaving disputes to be resolved at a diplomatic level (Reed 2011).
2. ***Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (OECD)***: Established by the Organisation for Economic Cooperation and development (OECD) in 1980, this broad set of principles are essentially a series of recommendations outlining good

practice in relation to data protection practices. Again, the primary aim here was to prevent restrictions on data flows between OECD member states. Similar to the Convention, the OECD has no formal authority. The OECD adopted a further declaration on Transborder Data Flows in 1985. The signatories indicated an intention to promote access to data/information and related services. They also sought transparency in regulations relating to information. Additionally, it set out to develop common approaches to managing trans-border data flows and to consider implications for other countries when dealing with data flows (Lloyd 2011).

Despite the aspirations and intent of both the Council and the OECD, the regulatory protection for personal privacy that emerged was not uniform across signatories. This occurred for a number of reasons, not least the fact that not all Council of Europe members adopted the legislation. A lack of clear definition of key terms led to inconsistencies when individual countries adopted the key principles into national legislation.

In an attempt to overcome these inconsistencies, the European Community in 1990 published a draft *Council Directive on the Protection of Individuals With Regards to the Processing of Personal Data and on the Free Movement of Such Data*. It is important to view this draft directive in its wider context. It represented one element of a shift to widening Europe into both an economic and political union, as embodied in the Treaty of European Union signed 1992 (Cate 1999).

The key piece of European data protection came in the form of the Directive 95/46/EC. Enacted in 1995, it gave EU Member States 3 years to bring national data protection legislation into line with the directive. The GDPR came into effect among member states in May 2018 and represents a major revision/updating of the key 1995 directive.

2.2.2 EU Data Protection: Core Principles

The EU Directive 95/46/EC became effective on 25th Oct 1998 and this data protection directive established information privacy as a basic human right.

“The EU Directive represents a unification of divergent European privacy policies, as it seeks to balance the privacy needs of the individual with the information needs of society” (Cate 1999, p431).

The underlying objective of the Directive was to advance the development and functionality of the single market (Murray 1998). Crucially, promoting the internal market required the Directive to balance two core principle objectives of; (1) Protecting the fundamental rights of an individuals right to privacy, and; (2) The prevention of obstacles to the free flow of information among Member States (Maxeiner 1995).

Directive 95/46/EC set out eight broad principles, these are:

1. **Purpose Limitation Principle:** Requires information only collected for specific and specified purpose and only used in ways compatible with those purposes. Information should be stored for no longer than is necessary for those purposes. Interestingly this principle also states that information not required for such specified purposes should not be collected.
2. **Data Quality Principle:** Requires that information be accurate and up-to-date.
3. **Data Security Principle:** Requires measures appropriate to risks be taken to protect data from improper use and disclosure as well as the security of data processing and transmission.
4. **Special Protection for ‘Sensitive’ Data Principle:** This imposes restrictions on the collection and processing of certain types of

information where data collected can identify a persons racial or ethnic origin, political opinions, religious or philosophical belief or concerning health or sexual orientation.

5. **Transparency Principle:** Requires data controllers to notify individuals when data is being collected about them; the purposed for which that data will be used; and the identity of the person who is collecting the information.
6. **Data Transfers Principle:** Restricts authorised collectors of personal information from transferring that information to third parties without the consent of the data subject. In the case of transborder transfers, the directive prohibits transfers outright to countries lacking what they call 'an adequate level of protection'.
7. **Independent Oversight Principle:** Requires an effective and independent supervisory body with the authority to audit data processing systems. This includes the ability to investigate complaints brought by individuals and the ability to enforce sanctions against data processors that fail to comply. The requirement for all individuals who engage in data collection to register with the relevant national supervisory body.
8. **Individual redress Principle:** Establishes that individuals have a right to access their personal information including the right to correct inaccurate information and the right to seek recourse against data controllers who do not comply with the directive.

(Loring 2002)

The EU Data Protection Directive established a high water mark in terms of regional and international legal protection in relation to information privacy (Cate 1999).

2.2.3 Interpreting the EU Directive: Working Party 29

Each member state appoints a Supervisory Authority who is responsible for ensuring compliance with the relevant provisions. In the UK this is the Information Commissioners Office (ICO). Article 29 of the Directive 95/46/EC created a '*Working Party on the Protection of Individuals with regards to the Processing of Personal Data*'. Often referred to as the Working Party 29, this group is composed of representatives from each of the member states data protection authorities. Its main purpose was to interpret key portions of the directive. For example, although Article 25 of the directive used the term 'adequate' in relation to data transfers yet no clear definition of this key term was included. The working party established a methodology to determine whether or not non-member states provided a sufficient level of data protection. This included setting out a number of procedural and substantive principles that must be met as a precondition to the finding of 'adequacy' (Loring 2002). There are also issues with regard to interpretation of the word 'relevant' in relation the Data Protection Directive.

The EU Privacy and Electronic Communications Directive (2002/58/EC), also known as the e-Privacy directive, translates the principle of the 'technologically neutral' Data Protection Directive into specific rules applied to the electronic communications sector.

The GDPR, published in 2015 and effective from 2018 supersedes the 95/46/EU directive. The GDPR incorporates many of core principles of the earlier directive while also attempting to address specific areas of concern to regulations as a result of more recent technological advances. This thesis examines the 'novel' new provision for the establishment of privacy seal accreditation schemes.

2.2.4 Privacy Regulation in the US

The US adopted a markedly different approach to the regulation of data protection compared to the EU. The EU Directive 95/46/EC can be categorised as a centralised, standard-setting approach that includes enforcement features. In contrast, the US regulatory environment is fragmented and narrowly targeted, addressing specific concerns/industries in reaction to perceived issues as they arise (Shaffer 2000). A core premise of the US approach is that openness of information advances privacy interests as it allows citizens to access and correct information held about them in an affordable manner. Any requirement to increase privacy protection will add costs to information flows, creating inefficiencies in the form of reduced productivity and higher prices for products and services and a loss of competitiveness for the macroeconomy (Cate 1999). Overall the US is a self-regulatory approach that endorses the core principles as set out by the OECD guidelines in relation to data protection (to which it is a signatory) and trans boarder flows.

The fragmented approach in the US means enforcement of data protection/information privacy is distributed and includes constitutional protections, common law privacy torts as well as state and federal legislation (Loring 2002).

In terms of constitutional protection, the First, Fourth, Fifth and Fourteenth Amendments of the US Constitution all contain limited protection in relation to an individuals privacy rights. However, most of the constitutional protection that does exist relates specifically to unwarranted government invasions of privacy. A landmark U.S. Supreme Court case of Whalen v. Roe in 1977 examined a New York statute requiring physicians to provide details relating to prescriptions to a centralised database. The Court recognised an individuals right to have their personal information remain private but also added,

“We are not aware of the threat to privacy implicit in the accumulation of vast amounts of personal information in vast data banks or other massive government files...” (US 1977).

Interestingly, the US Supreme court also ruled that no constitutional protections existed in the Fourth Amendment for personal information transmitted to their parties for commercial use.

A number of Federal Statutes also offer privacy protections. Here the approach has been reactive with legislation introduced on a piecemeal basis in response to specific issues that arise in an industry sector. Examples of such legislation include the *Fair Credit Reporting Act (FCRA) of 1970* which permits consumer credit agencies to release individuals credit information to businesses with a legitimate need for such information. *The Privacy Act 1974* regulates the collection, maintenance and dissemination of personal information by federal agencies. However, criticisms include the acts lack of an enforcement mechanism to ensure compliance. *The Electronic Communications Privacy Act (ECPA) of 1986* safeguards all electronic communications from unauthorised interception. *The Health Insurance Portability and Accountability Act (HIPAA) of 1996* regulated the use of private health records. *The Children's Online Privacy Protection Act (COPPA) of 1998* restricts the online collection of personal information for children under the age of thirteen.

The Gramm-Leach-Bliley Financial Modernization Act (GLBA) 1999 represents the most comprehensive US approach to privacy regulation. Some commentators' suggest US lawmakers were influenced by what the EU had done in relation to data protection in their directive. However, as with previous legislation, this act was limited in so far as it only applied to 'financial institutions'.

Some privacy regulation also exists in the form of State Statutes but these are limited by their application at a state only level. Despite this, some states have been active in creating industry-specific privacy legislation. Finally, the Common Law Tort of Privacy offers protection from individuals and the government in relation to privacy intrusions. However, it provides minimal protection for information privacy (Loring 2002).

As you might expect with a fragmented regulatory approach, responsibility for the enforcement of these rules in relation to privacy falls across a number of different agencies with the US Federal Trade Commission (FTC) and the US Department of Commerce as the primary enforcers. The FTC has issued a set of '*Fair Information Practice Principles*' or (FIPs) in relation to Data Protection but these can be seen as good practice principles reliant on the industry self-regulation and are not enforceable by law. However, other bodies do have powers in relation to specific acts. For example, the Securities and Exchange Commission (SEC) has enforcement powers in relation to the GLBA.

2.2.5 Managing Transnational Data/Information Flows

As we have seen, the EU and US take very different approaches in terms of how they address issues related to protecting information privacy. Through its Directives, the EU sets out clear principles and requirements for member states in terms of creating both overseeing and enforcement agencies. In contrast, the US takes adopts the position that legislation is only required to deal with sector or industry specific issues as they arise.

Perhaps the one clear area of commonality in terms of overriding principles in relation to privacy is the importance they both place on the role free-flow of information impacts commerce. However, the philosophical approach that underpins this principle differs greatly between jurisdictions. The EU enacted the Data Protection Directive as they feared differing national legislation would obstruct the free flow of data. This contrasts with the US, where the overriding principle sees data protection as state interference that introduces transactional costs making markets less efficient. The differing regulatory approaches both place significant importance of ensuring information-flows, in recognition of the key role this plays in promoting economic growth. It was this common view that led to the creation of the 'Safe Harbor Principle' (later revised to Privacy Shield). Furthermore, as economic globalization advanced and the growth of global new media firms accelerated, the free flow of data/information between different jurisdictions came into even sharper focus.

2.2.6 'Safe Harbor' Principle

Article 25 of the 95/46/EC Directive extended the application of EC privacy laws beyond the jurisdiction of the EU Member States. It required all data from citizens of member states to be given the same protection as was afforded to it in the EU directive. Some commentators have interpreted this move by the EC as a deliberate attempt to establish the directive as a global standard for data protection (Cate 1999). Working Party 29 ruled that the US GLBA law was not stringent enough to comply with the EU 'adequacy' requirement. This led to the prospect of a transatlantic 'data war' and prompted the European Commission

and US Department of Commerce to meet and agree a set of conditions. These are generally referred to as 'safe harbor' principles/standards were established to avoid a potentially damaging trade conflict. US based companies who stated that adhered to the agreed standards would be considered as adhering to EU data protection requirements. Negotiations were not particularly problematic, as many of the core principles of the EC directive were compatible with the OECD Guidelines where the US had been a signatory.

The US-EU Safe Harbor Framework came into effect in 2000 (Lloyd 2011). Under the Safe Harbour compromise, US firms who transferred data collected in the EU to the US would self-certify that they adhere to the core principles set out in Directive 95/46/EC. This arrangement, although not without its critics, operated successfully for many years.

However, in 2015, the Safe Harbour arrangement was declared invalid by the European Court of Justice (ECJ). This ECJ decision came about as a result of a complaint taken against Facebook by Austrian law student Max Schrems. Based on leaked confidential documents from former US National Security Agency (NSA) employee Edward Snowden, Schrems successfully demonstrated that Facebook was in breach of EU data protection principles when transferring and processing data collected in the EU, to the US. In October 2015, the European Court of Justice (ECJ) ruled that Facebook indeed insufficiently protected Schrems data and therefore was in breach of Safe Harbor.

The ECJ decision forced the EU and US authorities to come up with a new legal framework to cover the transatlantic data flows. This new arrangement came into force in Feb 2016 under a framework known as the 'EU-US Privacy Shield'. However, this did not completely settle the matter and a number of court cases are currently pending in relation to the validity of the Privacy Shield arrangement. In April 2016, the Working Party 29 highlighted there major points of concern relating to deletion and collection of personal information, as well as concerns in relation to the Ombudsperson mechanisms used to resolve disputes. The GDPR has effectively superseded the Privacy Shield as the regulation goes beyond it in terms of its scope.

2.2.7 Regulation v's Free Market Approaches

Academic critiques of the approaches taken by the EU and US to data protection are broad in scope and often technical/detailed. However, a number of key themes emerge and these are discussed below.

Neoliberal/neoclassical economists generally dislike formal regulation and favour self-regulation of markets. We discuss this in more detail in later sections of this chapter. Their criticisms of the formal EU style regulatory driven approaches to data protection cite a number of problematic areas. Apart from the usual 'transactional costs' argument (i.e. compliance with regulation adds cost and this impacts on prices and competitiveness etc), they also argue that the length of time it takes governments formulate and enact legislation (for a particular technology) means the 'new' technology itself is often outdated or superseded in the meantime. There are also concerns that efforts to reign in technology through regulation will force firms to move to less privacy restrictive jurisdictions and this in turn will hurt jobs and economic growth. Furthermore, critics argue that any attempt to *control* technologies will force developers to design more complex systems that are more difficult to monitor (Beck 2001). Broadly speaking, critics of the EU style regulatory framework often favour a more free-market approach/solution.

However, the free-market approach also attracts criticism from privacy campaigners. Those in favour of regulation that being applied to specific industries cite how regulation can be used,

“to promote activities or results beneficial to the public good, to limit or halt activities or results harmful to the public good” (Picard 2002, p.70).

The telemarketing industry in the US provides a simple yet relevant example of how pure marketplace approaches can limit user choice. Until regulations were introduced that created a National Do Not Call Registry, individuals had no way of preventing unwanted and intrusive phone calls. The new rule requiring telemarketers to include Caller ID when making such calls, preserved individuals basic privacy 'right to be left alone' while allowing user to opt in or

out of such telemarketing calls. There are parallels between the telemarketing direct marketing debate and the current Internet tracking debate (Hoofnagle et al 2012). The absence of an enforceable option for Internet users to choose not to be tracked is limiting their choice.

Neither the US or EU approach to data regulation can be seen as efficient. Interestingly, despite their differing approaches data protection regulations, major 'white paper' reports by the EU, the FTC and the US Dept of Commerce all pointed to similar remarkably similar failings, acknowledging that their respective approaches are not effective, perhaps even obsolete. The failure to meet the challenges to privacy that come from new technologies may in fact damage commerce and economic growth going forward (Cate 2011). We will explore these reports in more detail in the following sections.

2.2.8 Regulatory Environment: Fit for Purpose?

Rohn (2018) suggests that media firms can gain a better understanding of their audiences by examining the data they collect on them. In this context,

“important concepts such as trust, authority and knowledge are very relevant”
(Rohn 2018, p.434).

This view is very much echoed by EU and US data protection enforcement authorities and gaining a more detailed understanding of the relationship between new media consumers and new media platform providers is a fertile ground for media management researchers.

In 2010, the European Commission published a consultation entitled, ‘*A comprehensive approach on personal Data Protection in the European Community*’. It set out a number of areas where it believed the existing 95/46/EC Data Directive was not providing adequate protection for individuals or business. In Jan 2012, following a period of consultation with a range of stakeholders, the Commission set out more detailed proposals for the reform of the EU's 1995 data protection rules. The reforms had the dual goal of strengthening online privacy rights and boosting Europe's digital economy. A new single law would address issues relating to how member states had interpreted the 95/46/EC rules differently as well as reducing what they describe as ‘costly administrative burdens’ and save business €2.3 billion a year (European Commission 2012).

To address these and other issues, the Commission had undertaken a review of the existing legal framework through a range of public consultations and studies. Overall they found the core principles of the Directive to be valid but identified several problematic areas they described as “posing specific challenges”. The proposed changes represented a convergence of views across consumer organisations, business association and Data Protection Authorities. These recognised that technology developments including social networking and cloud computing had resulted in an increasing level of risk to individuals’

personal data. The Commission states that it believes that existing data protection provisions are inadequate citing that,

“ways of collecting personal data have become increasingly elaborated and less detectable” (European Commission 2012, p.2).

The following sections set out some of the key challenges identified by regulators.

2.3 Privacy Policies

A privacy policy represents a binding legal agreement between the data subject and data controller. It operates on what is commonly referred to a ‘*notice and choice*’ principle. The wording of the policy gives ‘*notice*’ of what data the controller is collecting and how they can use/share the personal information gathered about the data subject. The service user or data subject then has the option to exercise their ‘*choice*’ by deciding whether to use the service or otherwise. Requirements from the US and EU data protection regulators means that almost all websites carry a privacy policy.

A new media firm’s privacy policy is generally accessed via a link/hyperlink on the hosts (data controller) website/app. In most cases the link appears in either in the footer of the webpage or from a link displayed during the registration or purchase process. Engagement in said processes constitutes acceptance of the privacy policy and permits the data controller to collect and use the data subjects personal data in the ways described in the policy. Most countries have a designated Data Protection Authority (DPA) to ensure that data collectors comply with its own stated privacy policy. For example, in the UK, DPA responsibility lies with the ICO.

Empirical research has long shown most privacy policies to be both time-consuming and difficult to read/understand. As a result, individuals (data subjects) rarely engage with privacy policies, even when they express concerns about information collection processes (Jensen, Potts and Jensen 2008;

Reidenberg, Breaux, Cranor, French, Grannis, Graves, Liu, McDonald, Norton and Ramanath, 2015; Vila, Greenstadt and Molnar 2004). Research from McDonald and Cranor (2008) estimated the economic 'costs' associated with reading privacy policies. They looked at privacy policies from 75 popular websites and found the average policy was 2,500 words long. Based on a reading speed of 250 words per minute, they calculated that it would take approx. 10 mins to read an average privacy policy. Next they estimated the average number of websites visited by web users to be 1,462 different websites a year. This means the average person would need to spend 40 mins a day reading privacy policies alone! (McDonald and Cranor 2008). Clearly this is unrealistic expectation and regulators are therefore seeking alternative mechanisms to aid data subjects in this area.

2.3.1 Machine Readable Privacy Policies

An early attempt to address the complex area of privacy policies (for online users) was developed by the World Wide Web Consortium (P3P) during the 'dotcom' boom in the early 2000's. The initiative came about in response to specific concerns from the FTC who thought that data subjects privacy concerns might constrain the development of online commerce. It operated as follows. Website owners placed a privacy policy on their website in machine readable format. P3P provided tools embedded in a web browser that allowed the data subject to set customized privacy requirements/preferences. When a user visited a website, the privacy requirements set by the data subject were compared to the host website (data controller) with any differences between the two highlighted. The data subject could then decide whether they wanted to continue with in their interaction with the website. This process did not include the use of icons. While the P3P initiative retained the '*notice and choice*' principle at its core, the now machine readable privacy policy was successful in terms of vastly reducing the 'economic cost' of reading the policy.

Overall the P3P initiative enjoyed limited success. It failed to become more widely adopted due to criticisms of the interface design and the technical complexity in configuring it further limited its adoption (Hochheiser 2002).

Furthermore, the lack of clear incentives for data collectors to adopt the P3P initiative mean also contributed to the failure to achieve widespread adoption.

2.3.2 Regulatory Changes Proposed

The Commission's report set out a number of overall 'key objectives' in areas where data protection could be improved. These effectively reiterated the founding principles of EU data protection i.e. 1) strengthening individuals' rights and 2) enhancing the internal market dimension. We will look at these in more detail.

The 2012 report set out its desire to *Ensuring appropriate protection for individuals in all circumstances*. It identified how the definition of 'personal data' in the existing 95/46 Directive was not clear. It also noted how in data mining and data aggregation innovations meant that aggregating data across different systems/databases could identify data subjects. The report also made the case for the expansion of the current definition of what constitutes 'sensitive data' citing in particular areas like genetic data and location data is not covered under data protection.

Transparency and control are other key areas addressed in the report. The Commission accepted that privacy notices are unclear and do not always comply with existing rules. As such, they actually reduce transparency and control for data subjects. In order to enhance control, the report sets out aspirations in terms of 'right to be forgotten' and 'data portability' (recognised as particularly challenging in the new media environment).

The report formally acknowledged that there are differences in how individual Member States interpret how consent is attained (by data controllers with respect to processing of an individual's personal data). The acceptance/interpretation of implied consent was viewed as particularly problematic area and the Commission called for clarity on the rules of consent.

There was also a clear recognition that existing sanctions for serious data protection violations need strengthening and suggests that explicit criminal sanctions may also be needed. This is especially true for large new media firms where the maximum fine applicable represents only a tiny fraction of revenues.

The Commission set out areas where data protection regulations could help to promote the internal market dimension. It acknowledges that the 95/46 Directive has been incorrectly implemented in some member states. This has resulted in greater restrictions for some businesses in relation to the free flow of data. The new regulation would seek to provide greater legal certainty and a level playing field for data controllers. It would also aim to simplify the 'administrative burden' placed on data controllers under the current system.

The Commission also proposed that any new regulation included a 'Privacy by Design' principle. This is where technology infrastructure (software and hardware) would include the principle of privacy in at the early design stage of their systems architecture.

The report also called for the establishment of a EU privacy seal certification scheme that could be offered for products and services that meet the established privacy standards.

The eventual outcome of this process was the publication in 2016 of the GDPR. Member states were given an approx. 2-year implementation period and the new regulation into effect in May 2018.

The Commission highlighted other significant issues in relation to '*Clarifying and simplifying the rules for international data transfers*' and '*promoting universal principles*'. However, perhaps the most significant aspect of what the Commission set out is the sheer number and range of areas in which the existing Data Protection legislation was seen to be failing. The report concludes with broad aspirations,

"the challenge to legislators is to establish a legislative framework that will stand the test of time" (European Commission 2010, p.18).

This reflects the core challenge that regulators face when attempting to assemble the new regulation - crafting a regulation that will be able to keep pace with the rate of technological development.

Identifying areas that need improving is one thing, but implementing rules/regulations that address these current concerns while may well prove to be extremely difficult to achieve. The section below provides an example of some of the challenges faced by regulators when attempting to clarify how member states interpret the issues of 'consent' in the 95/46 directive.

2.3.3 Regulation Complexity: The EU Cookie Law

EU Directive 2009/136/EC (or 'Cookie Law') set out to bring clarity on one of the key principles of the 95/46 Directive, the issue of consent. Most online services use 'cookies' on their websites. Cookies are small text files that a data controller places in an individual's web browser when they visit their website. From a new media user perspective it means that when you return to a website you have visited before, the website may 'remember' you. This means that usernames/passwords as well as other preferences can be personalized to what you configured on previous visits. The 2009 directive sought to bring clarity in relation to how data controllers were permitted to use such cookies. Commonly referred to as the 'cookie law', regulators wanted to clarify whether *implied* or *explicit* consent was required by operators (data controllers) when using 'cookies'.

The Information Commissioners Office (ICO) in the UK summarised the issue here:

"The rules in this area are essentially designed to protect the privacy of internet users – even where the information being collected about them is not directly personally identifiable. The changes to the Directive in 2009 were prompted in part by concerns about online tracking of individuals and the use of spyware. These are not rules designed to restrict the use of particular technologies as such, they are intended to prevent information being stored on people's computers, and used to recognise them via the device they are using, without their knowledge and agreement." (ICO 2012, p.2).

The ICO stated that from 26th May 2012 implied consent would not be sufficient to meet EU data protection regulations. However, just 48 hours before the deadline, the ICO altered its advice/guidance stating that, "implied consent is

certainly a valid form of consent” (ICO 2012, p.7). This late change essentially shifted the responsibility for consent from the website operator to the user, thus shifting the burden for the user to opt-out rather than forcing the operator to get users to opt-in.

On one hand, the late guidance issued by the ICO can be viewed as a small semantic change in language. However, it captures the challenges faced by member states when implementing the regulation. This time in a *relatively* straightforward issue of consent, we see the same organisation, the ICO, change its advice/interpretation in the final hours before the directive becomes effective in law. In this example, it would appear that objections from UK businesses and website operators may have contributed to the ICO late change to its guidance. As the implementation deadline approached, there was pressure from the technology industry to move away from its explicit consent requirement. A report from KPMG had suggested that 95% of UK companies had not complied with the legislation and many risked going out of business should the ICO apply the max permitted £500,000 fine (Arthur 2012).

Overall, this example suggests that when it comes to the crunch, the promotion of commerce and maintaining the free flow of information trumps individual’s personal privacy protection. Similar issues have arisen in the US where proposals from the Obama administration sought to implement a clear and universal “Do Not Track” principle. However, successful lobbying by the new media industry saw Facebook’s “like” and Google’s “+1” buttons/features emerge exempt from Do Not Track principle (Hoofnagle et al. 2012).

The contents of the ‘Cookie Law’ were originally set out in the e-Privacy Directive in 2003 and these did not come into play in the UK until almost a decade later. Additionally, by the time the new rules came into force, technological developments had rendered the rules somewhat irrelevant. Research published in the Harvard Law & Policy Review journal,

“empirically demonstrate that advertisers are making it impossible to avoid online tracking” (Hoofnagle et al. 2012).

They show how tracking technologies have moved on from the use of 'cookies' that can be deleted by users to 'flash cookies' that can track individuals irrespective of whether they wish to be tracked or not.

The 'cookie law' example captures many of the challenges regulators face when attempting to make (even relatively minor) changes to data protection regulations. Firstly, it can take a long time to implement. In this case almost a decade from inception to it actually becoming law (i.e. 2003-2012). Secondly, the rate of technical change means that by the time the rule becomes effective, technological developments may render it obsolete i.e. the use flash cookies. Thirdly, in this case it would appear that powerful lobby groups were able influence the advice of the national regulator, as it did with the ICO in the UK.

2.3.4 Adequate Provisions – Poor Enforcement?

The UK legal academic Davies (2001) believes that the founding principles of modern data protection laws are essentially sound. He argues that existing legislative provisions offer substantial privacy protection for individuals, if enforced. He cites three key factors that contribute to this absence of enforcement.

Firstly, he argues governments regularly ensure their most vital areas of their function are exempt from privacy law, citing security or other public interest reasons. As a result, communication and information infrastructures are established with what he calls 'surveillance by design' as a core component.

Secondly, he believes individuals are overwhelmed by the complexity and processes required to enforce the protection for their privacy. The campaigning group Privacy International has estimated that personal data for the average citizen in the developed world is located on at least 400 databases. They estimate that gaining access to this data would take more than 40 working days to prepare, administrate and to analyse and that this is simply not practical for data controllers or individuals.

Finally, he describes data protection regulators as having been, “corrupted and compromised through timidity and neglect” (Davies 2001, p289). Here he is essentially accusing regulators of not enforcing the existing rules. Davies does acknowledge that some may interpret his views as ‘extreme’ and/or alarmist. However, he defends his position pointing out that once a fundamental right has been established, the rigorous protection of these rights is in fact a conservative notion and this it is the transgressor who becomes radical. He believes both the government and the private sector conveniently invert this notion.

Although one could argue that Davis views are somewhat dated, some of the points he raises may help inform future debates/regulations. For example, will it really be possible to distinguish one software application from another in terms of features that enhance a notion of ‘surveillance by design’ as described by Davies from something that includes ‘privacy by design’ as called for by regulators?

2.3.5 Differing Regulatory Approaches – Common Issues

As we have discussed, the founding principles of EU and US data protection regulations are based on distinctly differing founding philosophical approaches. One clear area of commonality is the key importance they place to the free flow of information. Although the philosophical approach that drives the principle differ greatly, both place significant importance on maintaining free information-flows as a core driver of economic growth and prosperity. It was this common view that led to the creation of the ‘Safe Harbor Principle’ (and its more recent reincarnation as the ‘Privacy Shield’).

Privacy campaigners call for tighter data protection regulation specifically to deal with these advances in technology. New types of databases can now store information across thousands of machines and process queries on trillions of records in just a few seconds (Varian 2014). Data protection regulators on both sides of the Atlantic recognize this and both sides agree that data subjects need greater transparency with regards to these practices.

Regulators in both jurisdictions have identified a number of areas where technological developments meant existing data protection regulations are failing to protect producers and consumers. We will examine these in more detail in the following sections.

2.4 Changes to EU Data Protection: Overview

The European Commission set out a comprehensive list of specific areas where data protection regulation requires enhancing. Addressing all of these is beyond the scope of this document and therefore the discussion will be limited to areas most relevant to the research being undertaken here. After a series of consultations with 'individuals and organisations' (i.e. a range of stakeholders) the report sets out what it is trying to achieve, again this is firmly in the context of the core principles of European data protection, namely:

- 1) Clarity in relation to how existing data protection principles apply to new technologies to ensure individuals personal privacy is protected.
- 2) The request particularly from multinational companies to reduce the administrative burden and increase legal certainty in order to create a level playing field for economic operators.

'Strengthening individuals' rights' is set out as a key objective to ensure appropriate protection for individuals' personal data. Personal data is defined as,

"all information relating to an identified or identifiable person, either directly or indirectly." (European Commission 2010, p.5).

The Commission identify three specific enablers of enhanced personal data protection. These are 1) transparency, 2) sensitive data categories and 3) control. We will examine these in more detail.

It is also useful at this to also set out how the GDPR defines personal data, it is:

“any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;” (GDPR 2018 Article 4(1)).

2.4.1 Transparency

The European Commission describes transparency as,

“a fundamental condition for enabling individuals to exercise control over their own data and to ensure effective protection of personal data. It is therefore essential that individuals are well and clearly informed, in a transparent way, by data controllers about how and by whom their data are collected and processed, for what reasons, for how long and what their rights are if they want to access, rectify or delete their data” (European Commission 2010, p7).

Regulators describe how online behavioural advertising makes transparency around who is collecting personal data and for what purpose, difficult for individuals to understand. Privacy notices, particularly in online environments, are described as non-transparent and not in compliance with existing rules. It sets out the basic requirements for transparency as,

“information must be easily accessible and easy to understand, and that clear and plain language is used” (EC 2010 p.7).

2.4.2 Sensitive Data

Advances in technology led to calls from regulators to review the list of categories categorized as sensitive data. In the 95/46 directive, sensitive data categories’ related to any data revealing, 1) racial or ethnic origin, 2) political opinions, 3) religious or philosophical beliefs, 4) trade-union membership, and the processing of data concerning 5) health or 6) sexual orientation. These

sensitive categories of data were initially set out in the ECHR in 1950 and require updating to reflect advances in technology over the last 60+ years.

The GDPR identifies special categories of personal data (or sensitive data) as,

“Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation shall be prohibited”. (GDPR, Article 9(1))

The GDPR adopts all six sensitive data categories as defined in the 95/46 directives. It also adds two additional categories to the list, 7) genetic data and 8) some biometric data. The regulation requires sensitive personal data to be treated differently from other categories of personal data i.e. it requires explicit consent from data subjects. It also has additional security requirements in terms of how the data is stored.

Processing of data is defined in the GDPR as,

“any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;” (GDPR, 2018 Article 4(2)).

2.4.3 Control

The Commission also set out the need for individuals to have, “effective control over their own data” (European Commission 2010, p8). They describe how online social networking poses particular difficulties and acknowledges evidence of circumstances where individuals have been impeded in exercising their rights in this area. The Commission states that it has,

“Received various queries from individuals who have not always been able to retrieve personal data from online service providers, such as their pictures, and

who have therefore been impeded in exercising their rights of access, rectification and deletion.” (European Commission 2010, p.8).

The Commission believe such rights should be clarified, made more explicit and possibly strengthened. It also sets out a ‘right to be forgotten’ principle for individuals when the data storage period has expired and/or consent has been withdrawn.

The GDPR is a comprehensive regulation and a discussion on all of its provisions is beyond the scope of this document.

2.4.4 Privacy Seals

As discussed previously, Recital 100 of the GDPR clearly calls for the establishment of a third-party accredited Privacy Seal certification mechanism to allow media firms to demonstrate compliance with the relevant regulation. EuroPriSE is one such EU based privacy seal accreditation scheme operator.

In their published ‘Cert Readiness Check’ document (available on their website), they set out the specific requirements the applicant must meet in order to be issued with one of their Privacy Seals. Many of the criteria direct reference the relevant EU Directive Article that the compliant is required to meet. For example, in relation to Online Behavioural Advertising (OBA),

“In WP 171, p. 11, the Article 29 Working Party considered that while OBA providers are responsible for the requirements of Article 5(3) of the ePrivacy Directive being met, publishers (website owners) have a certain limited responsibility for the data processing that results from the use of an OBA service.” (EuroPriSE, p.9).

EuroPriSE sets out the Transparency and Opt-Out criteria each website operator must meet in order to comply with the requirements of regulators.

The criteria that privacy seal operators will use to assess applications will reflect the specific articles from the GDPR. Firms who apply for Privacy Seals will need to demonstrate that they meet the criteria in the regulation.

2.4.5 US Federal Trade Commission (FTC) - Data Protection Proposals: Overview

Echoing their counterparts in the EU, US data protection regulators also acknowledge that existing provisions are not sufficient. In 2012, the then chairman of the US FTC Jon Leibowitz commented,

"right now, it is almost impossible to figure out which apps collect data and what they do with it." (Arthur 2012).

There have even been suggestions that if the online advertising industry in particular fails to adequately address the issue of consent in relation to how they collect, share and process data, the industry may find itself facing a Leveson type enquiry in the not too distant future (Kirwan 2012). An undermining of confidence in how personal information is collected and shared poses risks to the whole new media sector as well as overall economic growth. The following sections highlight some of the key areas where US data protection needs addressing.

Similar to EU regulators, the FTC identifies transparency as a key issue and that needs to be addressed. In relation to privacy policies it states,

"privacy notices should provide clear, comparable, and concise descriptions of a company's overall data practices. They should clearly articulate who is collecting consumer data, why they are collecting it, and how such data will be used" (FTC 2010, p.71).

Improving privacy policies is just one a series of measures can improve transparency. These are summarised in the table below.

Figure 2-2 FTC Data Protection Principles

GREATER TRANSPARENCY

Companies should increase the transparency of their data practices.

- Privacy notices should be clearer, shorter, and more standardized, to enable better comprehension and comparison of privacy practices.
- Companies should provide reasonable access to the consumer data they maintain; the extent of access should be proportionate to the sensitivity of the data and the nature of its use.
- Companies must provide prominent disclosures and obtain affirmative express consent before using consumer data in a materially different manner than claimed when the data was collected.
- All stakeholders should work to educate consumers about commercial data privacy practices.

(FTC 2010, p.IX)

As we can see from the table above, the FTC are calling for greater clarity in relation to what personal data is being collected and what this data may be used for. They later published a ‘recommendations’ report for businesses and policymakers outlining specific measures to improve transparency. This included a proposal for a food ‘nutrition labelling’ type approach with regards to online behavioural advertising. This includes a suggestion for standardising the format of privacy policies as well as outlining the potential role of icons in alerting users to personal data that may be used for (FTC 2012). However, it did not provide any detail how a ‘nutrition labelling’ approach would be implemented in practice.

2.4.6 US Department of Commerce: Data Protection Proposals

The executive summary of this report recognises the role that networked information technologies have played in transforming the US economy and social life over the last 15 years. It acknowledges a lag between what it describes as the ‘intensive’ use of personal information and the corresponding development of privacy regulations. The result of this lag,

“leaves consumers with a sense of insecurity about whether using new services will expose them to harm” (DoC 2010, p.148).

They call for the US government to recognize a full set of Fair Information Practice Principles (FIPPs) as the foundation for commercial data privacy. These principles should promote increased transparency,

“through simple notices, clearly articulated purposes for data collection, commitments to limit data uses to fulfill these purposes, and expanded use of robust audit systems to bolster accountability” (DoC 2010, p.150).

However, consistent with its free-market principles, the DoC clearly indicates its preference for a self-regulatory approach in this area, stating it,

“requires companies to recognize that different business models based on different personal data raise different privacy risks” (DoC 2010, p.18).

2.4.7 The ‘White House Report’ 2012

Based on the proposals/comments from both the FTC and the DoC, the 2012 ‘*White House Report*’, ‘*Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy*’, recognises privacy protections as, “critical to maintaining consumer trust in networked technologies” (White House 2012, p.95). The report declares the US privacy framework as ‘strong’ but lacking in two key elements, these are; 1) a clear definition of basic privacy principles that apply to the commercial world, and; 2) a commitment of all stakeholders to address consumer privacy issues that arise from advances of technologies and their associated business models.

Transparency and control are again identified as key enablers of privacy. The report states that,

“Consumers have a right to easily understandable and accessible information about privacy and security practices. At times and in places that are most useful to enabling consumers to gain an meaningful understanding of privacy risks and the ability to exercise Individual Control” (White House 2012, p.108).

The report calls for the establishing of a Consumer Privacy Bill of Rights to provide,

“general principles that afford companies discretion in how they implement them. This flexibility will help promote innovation. Flexibility will also encourage elective privacy protections by allowing companies, informed by input from consumers and other stakeholders, to address the privacy issues that are likely to be most important to their customers and users, rather than requiring companies to adhere to a single, rigid set of requirements.” (White House, 2012, p.98).

This discretion in terms of how it is implemented is consistent with the free-market philosophy the US has followed with regards to data protection regulation.

2.4.8 Summary

As we have seen, the development of electronic communications and in particular the rapid growth of the new media internet based services poses significant challenges for data protection provisions. The regulators reports show the EU and US authorities united in the acceptance that their respective data protection environments are not providing the required level of protection for data subjects or data collectors. Furthermore, they identify transparency and control as key elements in protecting individuals' privacy. This very much echoes the view of media scholars like Rohn (2018) who identify trust and authority as a key issue for 21st century media scholars. As well as providing clear definitions of key terms like 'personal data', EU and US regulators identify the importance of giving media consumers greater 'control' and 'transparency' over their personal data to help engender trust between the two parties.

The focus on transparency and control by regulators is consistent with academic definitions of privacy discussed earlier in this chapter. However, it is worth noting that individuals are only in a position to exercise meaningful control (i.e. what elements of personal data to provide to data controllers) of their personal data if they are provided with a transparent view of what their data can be used for and if/how it will be shared with third parties. In other words, they go hand in hand, transparency is a key enabler of control.

Formally, information transparency can be defined as, "the degree of visibility and accessibility of information" (Zhu 2002, p.93). Features that promote information transparency are those that give individuals access to the information that a firm has collected about them as well as how that information may or indeed will be used in the future (Awad and Krishnan 2003). Some commentators argue it is more than just the quantity of information made available to a consumer that is important. The 'quality of the interface' that makes their data accessible that is also important.

While transparency is important, the complexity around how this is communicated to users through privacy policies, terms and conditions or some

other means is major area of concern for regulators (Granados, Gupta and Kauffman, 2009).

EU regulators suggest that changes to user interfaces or use of icons could assist in improving information transparency in online environments. Article 100 of the GDPR explicitly calls for a certification scheme that uses marks and seals to demonstrate compliance. Privacy seals could potentially also meet US regulators desire to have,

“easily understandable and accessible information about privacy and security practices. At times and in places that are most useful to enabling consumers to gain an meaningful understanding of privacy risks and the ability to exercise Individual Control” (White House 2012, p.108).

However, none of these reports examined offer any empirical evidence to support the use of privacy seals/icons. Will the presence of an accredited Privacy Seal on a firms website provide the transparency and control assurances for individuals? The research outlined here will empirically test how use of iconography in the form of privacy seals impacts on personal information disclosure in online environments.

The following sections will examine examples of where icons/seals have been used in both online and offline environments when used to try and communicated messages to individuals in a standardized manner. We will also discuss existing empirical evidence that have examined how users behave when presented with privacy seals and/or similar iconography.

2.5 Iconography/ Privacy Seals

“Icons and labels both have a long history of helping communicating complex factual information in an easy-to-grasp way to consumers.” (Edwards and Able, 2014, p.2).

The General Data Protection Regulation (GDPR) facilitates the establishment of Privacy Seal certification mechanisms, specifically to deal with the issue of transparency. There is no specific requirement/obligation to for individual member states to establish their own certification mechanism. In the UK, the Information Commissioners Office (ICO), who has overall responsibility for ensuring UK data protection complies with EU regulatory requirements, is developing its own privacy seal accreditation scheme (Falmer 2015). The ICO stated on their website,

“A privacy seal is a ‘stamp of approval’ which demonstrates good privacy practice and high data protection compliance standards. It will work much like the British Standard Institute’s Kitemark symbol, which is displayed on numerous products and services within the UK to demonstrate quality and high standards.” (ICO 2016).

However, the 2016 referendum result in the UK where the UK voted to leave the EU has created uncertainty about the UK’s future relationship with the EU and this seems to have stalled the progress of ICO’s plans in this area.

While the EU’s explicit call for a data protection seal/mark certification mechanism to address online information privacy concerns is new to the technology sector, similar schemes, in both the offline and online work have been established before, with somewhat varying degrees of success. The following sections will provide a summary of such schemes.

2.5.1 Icons, Labels and Privacy Seals

In a data protection context, the primary function of ‘iconography’ (i.e. icons, labels, privacy seals etc) is rooted in the “notice and choice” principle. From a

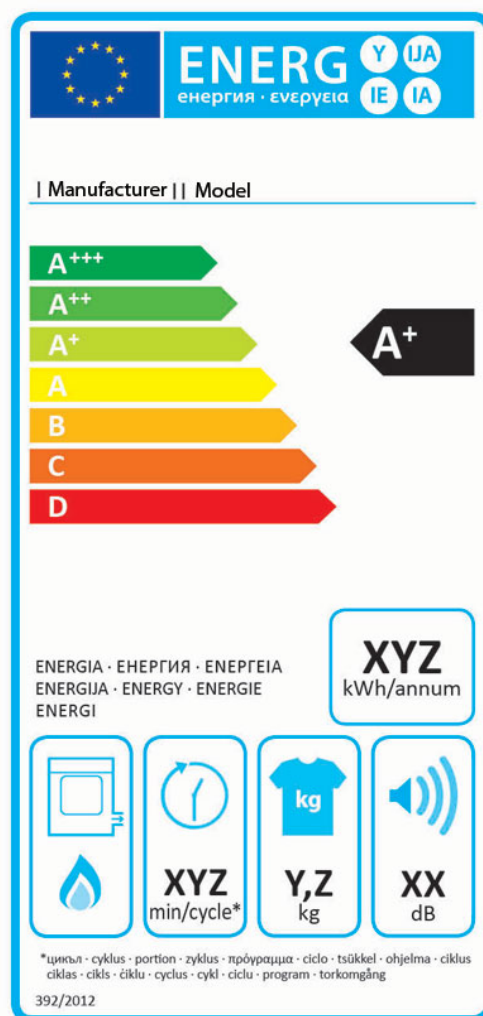
media consumer (data subject) perspective, the presence of iconography that demonstrates the compliance with relevant data protection regulations may offer reassurances with regards to how media firms (data controller) manage their personal data. As already discussed, data subjects struggle to read and fully understand privacy policies. The use of iconography has the potential to aid a data subjects decision-making in terms of whether to engage with said service provider (i.e. data collector) or not, without having to read the whole privacy policy. For data collectors, displaying iconography helps communicate key elements of their privacy policy in a more 'user friendly' form. Some media firms may see their use as a potential source of competitive advantage, as it allows data controllers to demonstrate their compliance with data protection regulations while also offering data subjects' greater transparency in terms of how their personal data is being collected/used/shared.

It is important to note that iconography does not give data subjects any additional control over their personal data collector per se. It acts as a visual representation of how aspects of a particular service operates i.e. what categories of personal data are collected, how it is stored and does it comply with the relevant data protection regulations. The following sections will explore the use iconography in both online and offline environments and offers some potential insights for how media firms may be able to build trust with media consumers.

2.5.2 Offline Iconography: Energy and Nutrition

The EU energy label system is a good example of how iconography is successfully used to aid consumer decision-making. In Sept 1992, Council Directive 92/75/EEC established a mandatory energy label system for all major household appliances. Standardized energy labels allowed consumers to see summary energy consumption and performance information specific to each appliance without having to read through an instruction manual. Below is an example of an energy label. As you can see, the label provides key information relating to the 1) energy efficiency rating; 2) estimated annual energy consumption; and 3) product specific information such as water consumption and/or noise levels, for each appliance.

Figure 2-3 Energy Label Example



1. Energy Efficiency Rating

A+++ is the most efficient, and D is the least efficient, based on the product's energy consumption.

2. Annual Energy Consumption

The annual energy consumption (in kWh per year) for each product is calculated using specific EU-defined criteria. Here, for tumble dryers, the figure is calculated based on the standard cotton program at full and half load.

3. Product-specific information

You'll also find images showing extra data related to the product, such as capacity, water consumption and noise levels.

With major household appliances accounting for 35% of residential end-use electricity consumption, energy labels have played a key role in both the diffusion of energy efficient appliances and helping to achieve overall energy saving targets (Bertoldi and Atanasiu 2007). Industry data shows that consumers are more likely to purchase appliances with higher efficient energy ratings. While some minor criticisms of energy labeling do exist, overall the scheme is viewed as a major success in terms of its impact on consumer purchasing behavior as well as the wider EU energy savings.

Food nutrition labels provide another useful example of offline iconography.

“Nutrition labelling is an attempt to provide consumers, at the point of purchase, with information about the nutrition content of individual food products, in order to enable consumers to choose nutritionally appropriate food.” (Grunert and Wills 2007, p.385).

In the US, the FTC believes food nutrition style labeling could play a useful role in communicating data protection practices (FTC 2012). In a major review of over 58 EU nutritional literature papers/reports, Grunert and Wills (2007) research concluded that consumers generally understood the link between food and health but that the degree of interest in this varies across situations and products. Their research showed that consumers were shown to generally like front of packet displays of nutritional information. The majority of participants in the study believed they understood the nutritional information presented to them and could recall it afterwards when prompted in some of the surveys/experiments. However, they also concluded that little is known in relation to how labeling information impacts on consumer’s overall wider dietary behavior/patterns.

The offline examples of iconography discussed above took relatively complex information sets and presented summary labels/icons to consumers in the form of a colour coded and somewhat standardized format. Research shows both systems were well received by consumers. However, in the case of the nutrition labels, researchers are unable to isolate the overall impact on the wider population dietary behavior. Indeed, as obesity rates continue to rise perhaps

this is evidence that the nutrition labels are effective at communicating a message but not at altering behavior on a wider level.

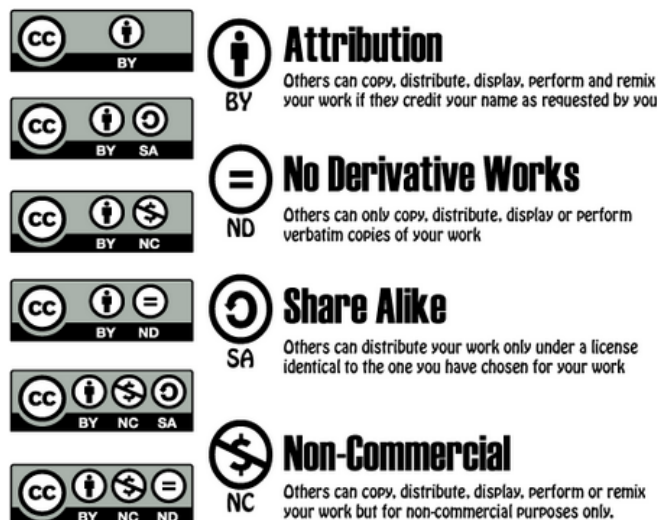
In the following sections will examine the use of iconography in online environments.

2.5.3 Online Iconography: Creative Commons Licences

In broad terms, individuals who create and distribute copyrighted works do so under an “all rights reserved” principle. This means that the creator retains the exclusive rights to reproduce/copy, adapt, distribute/publish, display or perform the work in public unless they have specifically given someone else such rights permissions.

The philosophy behind Creative Commons (CC) licenses is that it allows copyright owners to distribute their works with only “some rights reserved” (rather than all). Creating a CC license is user friendly and does not require extensive copyright legal knowledge. The image below sets out the different criteria for granting rights to use the works and these include; a) Attribution, this requires the accreditation of the original creator/author; b) No Derivative Works, the work can only be reproduced in its original format i.e. with no editing/alterations; c) Share Alike, others can distribute the original work but only under identical CC licenses; and, d) Non-Commercial, permits copy, editing and distribution of the original work if for non-commercial purposes only.

Figure 2-4 Creative Commons Licence



(Source: Edwards and Able 2014, p.12).

As we can see in the figure above, CC licenses have adopted a set of standardized icons.

“The icons appear clear and concise, and communicate otherwise complex intellectual property and copyright concepts in an easy to understand way.” (Edwards and Able 2014, p.10).

However, there has been no formal research to specifically assess the success of the icon approach adopted by CC.

The CC ‘tools’ operate as a globally recognized framework, developed in conjunction with legal experts in over 70 jurisdictions. Over 350 million CC-licensed works have been published by authors on the Internet (Creative Commons Factsheet 2017). While many CC licenses are sought for non-profit uses, a number of large commercial websites including Google and Flickr as well as a host of national governments use them, thus demonstrating a global acceptance.

The relative success of the CC license led to attempts to adopt a similar icon/graphical tool specifically for web users. The following sections highlight some of these attempts.

2.5.4 Online Privacy Seals

In an online context, privacy seals are used to,

“symbolically communicate a third-party authority designed to engender trust in the Web site’s information practices as stated in their privacy policy” (Rifon, La Rose and Choi 2005, p.341).

Privacy seals primarily serve two key functions. Firstly, they potentially raise the practices/standards of data collectors (who generally have to demonstrate compliance with the seal awarding body); and, secondly, to influence consumer (data subjects) perceptions (Miyazaki and Krishnamurthy 2002).

A number of organizations run their own online privacy seal schemes. TRUSTe is one such third party online privacy seal program and was founded in 1997 by

the Electronic Frontier Foundation (EFF) and CommerceNet Consortium. Website owners (i.e. data collectors) pay an annual fee to participate in the program and in return for this,

“Seal authorities provide a set of guidelines and a voluntary enforcement mechanism to assure that the site abides by its own privacy policy.” (Rifon et al. 2005, p.341).

For website users, the presence of the seal is considered to help build trust in the data collectors stated privacy practices. The seals credibility depends on both voluntary compliance as well as monitoring from the seal program administrators.

Figure 2-5 TRUSTe Privacy Seal



A number of major online brands participate in the TRUSTe system including Paypal, Apple, eBay and Oracle. However, the scheme is not without its critics. TRUSTe was itself found to have used a third party to track information on its website, breaching its own privacy standards! It also faced criticism for failing to take actions against Microsoft and RealNetworks on the basis that the deemed software glitches were responsible for data breaches rather than it being a deliberate act (Rifon et al, 2005). The results of the limited empirical research to examine assessing the impact privacy seals like the TRUSTe program will be discussed in greater detail in later sections of this document.

Figure 2-6 ePrivacy Seal



In the EU we have seen a small number of relatively 'new' privacy seal accreditation schemes emerge in recent years. These private firms include the EuroPriSe 'European Privacy Seal' and ePrivacy 'European seal for your privacy' accreditation scheme operators (see figures above/below). In broad terms both schemes operate in a similar way. Firms who wish to display a

privacy seal must be able to demonstrate to the accreditation body that the way they manage their software (website/apps etc) or products meet the EU data protection regulations requirements under GDPR.

Figure 2-7 EuroPriSe Privacy Seal



2.5.5 Privacy Bird

Privacy Bird is an initiative from the Worldwide Web Consortium (P3P).

Figure 2-8 Privacy Bird



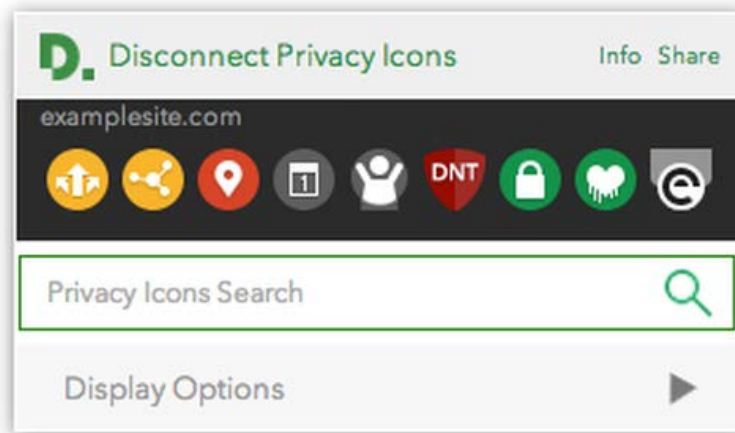
Users were required to set their privacy ‘framework’ settings in a web browser add-on and the host websites privacy policy was machine read when they visited a website. The privacy bird would then appear colored green, amber or red depending on how the user privacy requirements compared against the host websites stated privacy policy. However, Privacy Bird has not enjoyed widespread adoption.

2.5.6 Privacy Icons

A number of academic projects have created custom sets of privacy icons to use in their own research into how they may help users navigate the complex area of online privacy. The ‘Prime Life’ project from Stanford University in the US, who launched a set of ‘PrivIcons’ is one such example. However, to date, academic initiatives developed exclusively for the research objectives being undertaken in a specific study, have not been put to practice in a commercial ‘real world’ environment.

In 2014, TRUSTe in conjunction with open source software company Disconnect, launched a set of privacy icons (known as PIS, Privacy Icons Software). The intention here is to create a system for commercial rather than for academic use.

Figure 2-9 Disconnect Privacy Icons



The icons operate as a browser plug-in and are displayed for each website visited by the user. However, similar to issues with P3P, the icons purely report how the website operates with respect to its stated privacy policy and does not guarantee that the data collector is actually operating in accordance with the rules defined by relevant regulators.

The Open Law Laboratory at Carnegie Mellon University developed an initiative looking at Privacy Nutrition Labels (see example below). The idea is that for each website, a privacy nutrition label would detail what information is collected and for what purpose, while also detailing when information is going to be shares with partners on publically.

Figure 2-10 Privacy 'Nutrition' Style Label

Bell Group

information we collect	ways we use your information				information sharing	
	to provide service and maintain site	marketing	telemarketing	profiling	other companies	public forums
contact information		opt in			opt out	
cookies						
demographic information		opt in			opt out	
financial information						
health information						
preferences						
purchasing information		opt in			opt out	
social security number & gov't ID						
your activity on this site		opt in			opt out	
your location						

Access to your information
This site gives you access to your contact data and some of its other data identified with you

How to resolve privacy-related disputes with this site
Please email our customer service department

bell.com
5000 Forbes Avenue
Pittsburgh, PA 15213 United States
Phone: 800-555-5555
help@bell.com

There is no information available on privacy 'nutrition' labels being used commercially (i.e. outside of academic research).

2.5.7 Summary

The offline Standardized Energy Labels scheme is viewed as a success and there is potential for this approach to be replicated in online environments. The initiative saw complex technical descriptions of device performance displayed in easy to understand visual standardized labels that summarized key performance metrics. Standardized labels made it easy for consumers to compare the performance of different makes/models. Data shows that after their introduction, consumers generally opted for more energy efficient models. This enabled regulators to achieve the overall goal of reducing energy consumption and the related emissions from white good appliances.

With food nutrition labels the position is less clear. While at a micro level research shows that individuals understand the message being communicated,

there is little evidence of how it impacts at a macro level and rising obesity rates suggest that these labels are not achieving their desired behavioral impact.

In the online world, we see that the use of standardized icons/labels in the form of CC licenses has been widely adopted to communicate copyright restrictions.

In varying formats, privacy seals/icons have been developed to help communicate data protection practices by data controllers but the impact of this is less clear. While the TRUSTe initiative was widely adopted, it only indicated data controller's compliance with stated privacy policies rather than demonstrating compliance with a baseline standard or similar. The type of privacy seal accreditation system envisaged by EU regulators in the GDPR has only been attempted commercially relatively recently. There is no data available with regards to how widespread it has been adopted. More significantly, there is little or no empirical data showing what affect these seals have on personal information disclosure.

While many data subjects state concern about their online privacy, there is little evidence to confirm that privacy is a driving factor in terms of deciding what services users engage with. The lack of clear evidence offers little incentive for data controllers to implement major privacy enhancing programs or to participate in the type of privacy seal accreditation scheme as envisaged in the GDPR. In fact, one could argue the opposite, that by not collecting and/or sharing/selling personal information from their website visitors, data controllers may be putting themselves at a competitive disadvantage in terms of revenues they could derive from said data.

This is the information gap the research undertaken in this thesis will address. How do privacy seals that communicate data collectors compliance with relevant data protection regulations impact on the amount and types of data that data subjects are willing to disclose?

2.6 *Media Economics Approaches to Privacy*

The term media economics is often used to refer to the business and financial activities of firms operating across varying media industries. Robert Picard formally defines media economics as the field of study that examines how,

“media operators meet the informational and entertainment wants and needs of audiences, advertisers and societies with available resources” (Picard 1989, p.7).

Economics is an appropriate lens to examine media firms because,

“decisions taken by those who run media organisations are, to a greater or lesser extent, influenced by resource and financial issues. So, economics, as a discipline, is highly relevant to understanding how media firms and industries operate” (Doyle 2013, p.1).

Albarran expands the scope of media economics, describing how it,

“strengthens our understanding of the role and function of mass media in society.....by adding important dimensions regarding the structure, conduct, and performance of media firms and industries; the interplay of economics, policy, and regulation; and audience behaviors and preferences” (Albarran, 2004 p. 303).

How media markets are governed is a fundamental component of media economics. The regulatory environment impacts on how media operators meet the needs of advertisers, audiences and indeed wider society. We will now look at some of these key roles/features in more detail.

2.6.1 *Media Markets*

“Most media organizations are, in the simplest terms, in the business of connecting media texts with audiences, to earn money through direct purchase or pairing audiences with advertisers” (Arsenault 2017, p.12)

The economics model of traditional 'media' firms like newspapers and TV channels is based on the notion of 'two-sided' markets. The 'first' market can be thought of as relatively straightforward content delivery. Examples include newspaper/magazine copy/content or tv programmes and these are sold to consumers at point of sale or through subscriptions. An important 'secondary market' also exists where audiences attracted to this content earn advertising revenues based on content ratings and demographics (Picard 1989).

Developments in the media sector, primarily driven through advances in networked technology, have seen a fundamental change in how revenue streams are generated in the two-sided market. We now see a multitude of examples 'new media' firms, who provide digital services (search, social networking, email, messaging and mapping) where the 'first' market (i.e. content) is not physically sold to consumers and therefore earns almost no revenue. Indeed many of these firms provide their content/services for 'free'. Almost all income in this new media sector is generated through secondary audience markets i.e. advertising.

"In the era of social media, where extremely large volumes of personal data are being created, harvested and cross referenced for the benefit of these secondary markets, it is immediately obvious that the waste products of everyday social interaction are being subjected to an alchemical transformation.....A base material is being transformed into gold" (Athique 2018, p.63-64)

Although there are numerous firms operating in this industry, perhaps two of the best-known players include Facebook and Google. In terms of their dominance of digital advertising, in 2016 these two firms captured over 70% of digital advertising spend in the UK and in 2017 they had over 63% of the US digital ad spend (Koetsier 2018).

"Twitter, Facebook and YouTube focused on social media content delivery and organization. As they have matured, in an effort to monetize user-generated content, the companies have transformed from content driven enterprises into data-collection companies. Data that can be packaged and sold to advertisers and marketers is their main product". (Arsenault 2017, p.10)

There are trade-offs for both users and providers of 'free' online media applications. For example, in the case of social networking, users share personal data about their interests, hobbies and locations. In return, they have access to a media service that lets them communicate and interact with their 'network' of friends, share location information and share photos etc.

"The take-home message from existing studies on personal information collection and privacy is that consumers are willing to share personal information, as long as the perceived benefits (personalized offerings) exceed the perceived costs (privacy)". (Evens and Van Damme 2016, p.6)

As the service provider, new media firms earn revenues from online advertisers where users 'public' data allows them to create more efficient targeted advertising campaigns.

Many new media firms operations are enabled through networked technologies where privacy has become,

"an issue of strategic importance for companies operating in the information-centric networked global economy" (Awad & Krishnan, 2006, p 25).

As fixed costs associated with managing data have rapidly declined, we are seeing many new entrants mostly in the form of startups in this industry sector (Varian 2014). These developments bring risks for both individual as well as new media service providers and this has attracted the attention of both privacy campaigners and regulators.

Data from multiple databases can now be linked to create digital dossiers of individuals' activities and interests, often without their knowledge or explicit consent. As discussed earlier in this chapter, EU and US media industry regulators are concerned about the impact technological advances are having on privacy. The US Department of Commerce warns about the potential 'harms' and related impacts on new media consumers. They say,

“Commercial data privacy policy must address a continuum of risks to personal privacy, ranging from minor nuisances and unfair surprises, to disclosure of sensitive information in violation of individual rights.....In the aggregate, even the harms at the less severe end of this spectrum have significant adverse effects, because they undermine consumer trust in the Internet environment. Diminished trust, in turn, may cause consumers to hesitate before adopting new services and impede innovative and productive uses of new technologies” (DoC 2010, p.148).

Data protection regulations that fail to adequately protect risks to personal privacy may undermine consumer confidence and have wider negative implications for the new media industry.

2.6.2 Media Regulation

Governments play an important role in creating the legal frameworks necessary to make markets function effectively. Regulatory intervention in specific media markets can take the form of, 1) technical/structural and 2) behavioral regulation (Picard 2002). The rationale that supports the role of regulatory intervention includes the need to control the behaviors of media firms,

“when they do not serve important social, political, and economic needs of society” (Picard 2002, p.70).

Technical/structural regulation may come in the form of limiting the number of mobile phone operators to ensure stable operations or limiting monopoly power etc. Behavioral regulation of media firms can take the form of prohibiting certain practices as well as enforcing compliance with specific rules or practices. Proscriptive regulation may limit economic benefits available to media firms. For example, the prohibiting of alcohol and tobacco advertising on television cuts off potentially lucrative income streams for broadcasters from those industry sectors. Prescriptive regulation may take the form of requiring broadcasters to include specific mix of content in their schedules i.e. children's programs,

news/current affairs etc. In his seminal 1978 paper, George Akerlof makes the argument for government intervention on the grounds that it may increase the welfare of both buyers and sellers in the marketplace.

The rule requiring UK broadcasters to show a product placement symbol to broadcast content where advertisers have paid for specific products to appear is a recent example of behavioral regulation. Critics of the EU media regulatory framework argue there is too much focus on rules that extend competition within the sector while discouraging rules that protect wider conceptions of the public interest (Harcourt 2005).

2.6.3 Economic Approaches to Privacy

Privacy economics is primarily concerned with 1) understanding the costs and benefits that data subjects (or collectors) bear or enjoy when their personal information is either protected or shared and 2) understanding how market mechanisms, technology or public policy can assist in achieving a desirable balance between information revelation and privacy protection (Acquisti et al. 2016).

“Privacy as a subject of public policy refers to the possession and acquisition of knowledge about people and implicitly or explicitly also knowledge about associations” (Stigler 1980, p.624).

There are effectively 3 phases of the development of economics being used as an approach to privacy. They are:

- **Phase I:** First economic approaches to privacy appear in the late 1970's and represent a classical approach based on rational expectations.
- **Phase II:** Not much by way of research from this period. Economists focused on 'new' information technologies focussing on privacy 'calculus'.

- **Phase III:** The development of the commercial Internet (and related data gathering/mining technologies) from approx.. 2000 on saw a large increase in theoretical and empirical studies.

(Acquisti et al. 2016)

2.6.3.1 Privacy Economics - Phase I [1970's & 1980's]

The first economics approaches to privacy appear in literature from the late 1970's. Posner sets out three different meanings of privacy; 1) the concealment of information, 2) peace and quiet and 3) as a synonym for freedom and autonomy. He argues concealment of information is the most interesting definition from an economic perspective. Posner believes that privacy laws and statutes result in a reduction of market efficiency i.e. the additional costs lead to reduced wages and ultimately higher unemployment (Posner 1978).

Posner's views very much represent a free-market neoliberal approach to privacy. He relates the economics of privacy to the economics of information through the 'market' and asserts symmetry between 'selling' oneself and selling a product. A consumer purchasing a product is operating in a utility maximising mode. Posner argues that this also occurs when a consumer sells 'information'; the purchasing of a product and the selling of personal information are both utility maximising decisions. If fraud (i.e. the concealment of defects in a good) is bad in the latter context of selling a product, why is it acceptable in the former where it reduces the amount of information available and thus makes the market less efficient? Posner asks,

"Why would someone want to conceal a fact, except to mislead others in transacting with him?" (Posner 1981, p.408).

Using the jobs market as an example, he argues that a person protecting their personal information may negatively impact on the firms hiring decision by way of misrepresenting their background and work experience. This may negatively impact on the firms hiring decision and ultimately the profitability of a firm i.e. the opportunity cost of not having hired the correct person. Therefore,

regulations that assist to give a legal right to remove an individual's personal information from the 'marketplace' serves to,

"ultimately transfers the cost of that person's possibly negative traits onto other market participants" (Acquisti et al. 2016, p.10).

Stigler (1980) identifies a certain irony in the enormous interest in privacy as "evident in the public press and statute books" (Stigler 1980, p.623), citing the fact that the 'average' citizen lives with greater privacy by virtue of living in large cities and working in large organisations. Individuals also have the option of moving to another part of the country to shake off their past. He suggests the increase in privacy concerns is due to the growth of government and its access to larger quantities of data than ever before. He asserts,

"Technology has enormously changed the mechanics of gathering and disseminating information, but it is politics and economics that direct the uses of the machinery" (Stigler 1981, p.623).

Both Posner and Stigler represent a classical rational approach theory to decision-making where users maximise their 'utility' over time using all available information and expressing consistent preferences. They argue that government regulation through establishing rights to privacy distorts the market by reducing efficiencies. The finance industry argues this position by claiming that if they had more detailed information on loan applications personal data then they could lend more money at lower interest rates to low risk borrowers and that this would benefit the wider economy in terms of growth and stability (Acquisti et al. 2016).

However, there are competing theories as to the overall effect privacy regulations have on technology. The Classical approach sees privacy regulation as imposing costs on the exchange of information therefore inhibiting the diffusion of technology. Others argue that effective privacy regulation promotes the use of technology by providing reassurances to the public that their personal data will be safe (IBID).

2.6.3.2 Privacy Economics - Phase II: [1990's]

Privacy economics literature developed in the 1990's and this change was driven by the emergence of 'new' information technologies. Varian (1996) revisits the classic rational approach model and gives examples of when disclosure of certain pieces of information may promote efficiency by bringing buyers and sellers together. Although his work pre-dates Google's behavioural targeting technologies, the success of AdWords could be viewed as representing as a major step forward in advertising market efficiency. Information about users' preferences gleaned from their search terms and viewing habits helps advertisers to target audiences more efficiently. Market efficiency leads to reduced prices, more demand, lower unemployment and ultimately higher levels of economics growth and prosperity.

However, Varian identifies examples where information disclosure may not promote market efficiency. A seller knowing how much a buyer is willing to pay may in fact work to reduce efficiency by creating opportunities for price discrimination. Other challenges to the rational approach are also given. If AIDS test results were to be made publically available this may act as a disincentive to taking the test. A reduction the number of people who are tested may in fact contribute to the spread of the virus resulting in wider economic and social costs (Varian 1996). He also addresses how technology is facilitating person/private information to be re-sold and suggests that regulation is required to deal with this area. Arguably data protection legislation that appeared in both UK and EU at around this time was introduced to deal with the types of concerns these issues raised.

Noam (1997) also addresses the privacy issues that new technologies portend and discusses how encryption may be used to prevent personal information from being disclosed. Noam puts forward the notion of 'markets for privacy' and argues that encryption tools will allow consumers to keep personal information private unless they choose to sell it (Noam 1997). However, recent revelations in relation to the operations of the UK and US national security agencies by whistle-blower Edward Snowden can perhaps be seen a blow to the encryption argument. Papers show that encrypted data was specifically targeted by

security agencies (Greenberg 2013), perhaps reinforcing early neoclassical arguments of excessive private behaviour signals that the individual has something to hide?

The 'Phase II' development of privacy economics literature represents an important challenge to the free market (i.e. market efficiency) approach put forward by Neoclassical economists. Here we see arguments and examples where access to individuals personal data may promote price discrimination or lead to other undesirable outcomes that inhibit market efficiency,

“the ability to track buyers makes it possible for sellers to implicitly collude.”
(Acquisti et al. 2016, p.19).

2.6.3.3 Privacy Economics - Phase III: [2000's - present]

The development of the commercial Internet (and related data gathering/mining technologies) from circa 2000 has led to large increase in the theoretical and empirical studies exploring economic approaches to privacy. Since 2000, a major departure from the traditional rational choice theories has emerged. This work encompasses pioneering work from Herbert Simon's contribution in terms of *bounded* rationality. Rational choice theory as related to personal information disclosure argues that individuals make 1) sensible and consistent trade-offs between privacy and other concerns and 2) engage in “disclosure management” providing information when they expect a net benefit (John, Acquisti and Loewenstein 2010).

Addressing rational choice theory, Simon argues,

“Rationality denotes a style of behavior that is appropriate to the achievement of given goals....Theories of rational behavior may be normative or descriptive – that is, they may prescribe how people or organisations should behave in order to achieve certain goals under certain conditions, or they may purport to describe how people or organisations do, in fact, behave” (Simon 1972, p.361).

Bounded Rationality challenges this classical rational approach theory and focuses on the inability of an individual to calculate all the potential parameters/outcomes when specific choices presented.

“The concept of bounded rationality has its roots in H. A. Simon’s attempt to construct a more realistic theory of human economic decision making” (Selten, 1999, p.5).

Simon (1986) cites empirical research that examined the purchase of flood damage insurance as an example of where rational choice theory does not hold (Kunreuther, Ginsberg, Miller, Sagi, Slovic, Borkan, and Katz, 1978). Neoclassical theory would predict the decision on whether to buy flood insurance (or not) is dependent on the buyers expectation that damages received in the event of a flood will be greater than the insurance premium. However, research findings suggest that decisions to purchase insurance premiums are mainly based on whether individuals (or close associates of those individuals) have had experience of flood damage – independent of any cost/benefit ratio analysis. Utility maximisation is therefore not sufficient for deducing who will buy insurance but,

“the process that puts the item on the decision agenda - is the important thing” (Simon 1986, p.216).

Studies have looked at this issue in relation to decision-making in privacy sensitive situations (Acquisti 2004). Research outputs show how specific framing of questions or scenarios may influence our choice of options. Similarly, heuristics (i.e. acting intuitively) may replace rational searches for the best alternative and as such biases and other anomalies affect the way we address alternative options etc. Research outputs from Acquisti and Grossklags (2007) argue,

“Behavioral economics studies how individual, social, cognitive and emotional biases influence economics decisions.... Predominantly based on neoclassical models of economics behaviour but aims to integrate rational choice theory with convincing evidence from individual, cognitive, and social psychology.

Behavioral economic models often abandon some of the tenets of rational choice theory: that agents possess consistent preferences between alternatives, chose the utility maximizing option, discount future events consistently, and act upon complete information or known probability distributions for all possible events” (Acquisti and Grossklags 2007, p.6).

Kahneman (2003) describes bounded rationality as an architecture that moves away from the rational model where ‘agents’ don’t necessarily reason poorly but in fact often act intuitively. Behaviour is,

“not guided by what they are able to compute, but by what they happen to see at a given moment” (Kahneman 2003, p.1469).

He also identifies emotion is an important consideration in decision-making.

Behavioural economics builds on some of the key themes addressed in bounded rationality. While it will not provide a set of rules or lessons in how to avoid all risks/mistakes in behavioural economics, it can assist privacy research through providing a better understanding of privacy decision-making and behaviour. One of the main challenges identified in privacy research is incomplete information (e.g. information asymmetries and/or concealment), framing, heuristics and biases – all these elements play interdependent roles while not always being present (Acquisti and Grossklags 2007). Behavioural research in economics draws on findings from disciplines including psychology and cognitive science in an attempt to attain a better understanding of human behaviour as they relate to economics issues.

2.6.4 Privacy, Advertising and Electronic Commerce

“Personal data is considered a new currency in many digital business models” (Evens and Van Damme 2016, p.12)

Website operators, data aggregators and digital advertising networks can together track users behaviour across multiple online services using a variety of tracking technologies such as cookies, web bugs and flash cookies. While many

data subjects are often unaware they are being tracked, advertisers have developed new methods of tracking such as flash cookies to counteract sophisticated consumers attempts to avoid being tracked (Hoofnagle et al. 2012).

“Online advertising is perhaps the most common example of how firms use the large amounts of data that they collect about users. The greater availability of personally identifiable data on the Internet in terms of scope, quantity, and the precision with which firms can target specific users challenges the traditional distinction between personal selling and remote communication.” (Acquisti et al. 2016, p.24).

Online Behavioural Advertising (OBA) technologies also allow advertisers to target specific subsets of consumers based on their browsing and/or purchasing history in a way not possible by traditional advertising methods. The detailed measurement metrics available in OBA allows advertisers to make continuous refinements/improvement to their campaigns.

Regulators have attempted to restrict the ability of data collectors to gather personally identifiable information and the EU’s so-called ‘cookie directive’ (i.e. requirement for explicit consent as discussed earlier) is perhaps a good example of this. However, issues with interpretation of the provision led to some member states not requiring explicit consent (requiring consumers to opt-in) and only requiring implied consent instead. We see a similar call from the FTC with their 2012 Do-Not-Track proposal. However, again data subjects are required to opt-out rather than opt-in. One of the aims of the GDPR was to address some of these inconsistencies of implementation among member states.

A 2012 study undertaken at MIT provides empirical analysis in relation to how the EC 2002 e-Privacy Directive impacted on the effectiveness of banner advertising. The Directive had made it illegal for websites to use ‘web bugs’ (different to cookies because they are invisible to users) and other hidden so called ‘spyware’ technologies without users’ consent. While the findings primary focus on the negative effects of the regulation (i.e. neoclassical inefficiency arguments), it can be seen as an example for how a EU regulation can be

effective in limiting how user-browsing data is shared across applications without users knowledge (Goldfarb and Tucker 2011).

2.6.5 ‘Nothing to hide nothing to fear’

When discussing privacy issues, an argument regularly put forward is that no privacy problem exists if an individual has “nothing to hide” (Solove 2007). Comments made by the then Google (Alphabet) CEO Eric Schmidt in 2009 very much echo this view. In an interview relating to the Google search engine product, Mr Schmidt was asked if users should be prepared to share information with Google as a “trusted friend”. Schmidt replied to this by saying if you are doing something you do not want others to know about, then perhaps you shouldn’t be doing it in the first place (Gelles et al. 2009).

In a related argument, Judge Richard Posner contends,

“people today decry lack of privacy, what they want, I think, is mainly something quite different from seclusion: they want more power to conceal information about themselves that others might use to their disadvantage.” (Posner 1981, p.406).

This position is a very much from the Neoclassical school of thought, equating an individual’s privacy to a product sold in the marketplace. If we are not permitted to conceal information about a product or service defect then why should the law permit persons to conceal discreditable information about them?

The ‘nothing to hide’ argument can be problematic to dismiss in the context of surveillance by security agencies, where the direct implication is that those who object to such monitoring are involved in criminal or illicit activities and therefore do have something to hide/fear. The primary problem with the ‘nothing to hide’ argument is the underlying assumption that privacy is about hiding bad things. This is not the case,

“Even surveillance of legal activities can inhibit people from engaging in them.”
(Solove 2007, p.745).

Solove identifies two key problems with surveillance, 1) 'aggregation', where smaller pieces of innocuous data are combined to reveal a lot more about an individual's interests and activities, and 2) 'secondary use', where data collected for one use is used in the future for an unrelated purpose. The potential future uses of data are vast and difficult to know/regulate once the data has been collected.

2.6.6 Information Asymmetry

Both EU and US regulators see information asymmetry as a key issue that needs to be addressed to protect privacy for data subjects. Information asymmetry,

“refers to situations, in which some agent in a trade possesses information while other agents involved in the same trade do not.” (World Bank 2003).

As such, information asymmetry creates an information imbalance between data subjects and data controllers and regulators seen to think that the balance is weighted in favour of the data controllers (i.e. new media firms). The imbalance has the potential to have much wider macro effects on the new media industry.

“It is well recognized that informational asymmetries may, at times, result in market failure” (Doyle 2018, p.61)

We can see information asymmetry issues at play in the context of privacy decision-making. Most data subjects lack awareness of how data controllers could potentially use and/or share their personal information and this may result in data subjects engaging in more risky behaviour from a privacy perspective (Tsai, Egelman, Cranor and Acquisti 2011). Website privacy policies represent the clearest attempt to address this issue of information asymmetry (Milne and Culnan 2002). This 'notice and choice' approach has been widely and is indeed a legal requirement for many new media services (i.e. data controllers). In principle, privacy policies fill the information gap between data subject and data

collector by setting out what data is collected and how it is used/shared. However, as discussed, issues with the length and complexity of privacy policies makes them difficult and time consuming to read (Jensen and Potts 2004) resulting in people rarely reading them (Jensen et al 2005). Research also shows there is a considerable economic cost associated with reading the privacy policies of all the website and applications an average user engages with (McDonald and Cranor 2009). Even when users do take the time to read privacy policies, they often make mistaken assumptions about the meaning of the privacy policy. There are also issues with interpretation of what a privacy policy actually represents. Research from the US shows that many people incorrectly believe that the mere presence of a hyperlink to privacy policy means their personal data is protected (Turow, Hoofnagle, Mulligan and Good 2007).

In 2012, the White House Report set out some of the issues that arise in relation information asymmetry. The report notes,

“individuals who actively share information with their friends, family, colleagues, and the general public through websites and online social networking sites may not be aware of the ways those services, third parties, and their own associates may use information about them. Unauthorized disclosure of sensitive information can violate individual rights, cause injury or discrimination based on sensitive personal attributes, lead to actions and decisions taken in response to misleading or inaccurate information, and contribute to costly and potentially life-disrupting identity theft.” (White House 2012, p.100-101)

A 2013 survey by the Pew Research Centre found that 68% of US adults believe that the existing regulatory framework does not sufficiently protect their online privacy (Rainie, Kiesler, Kang, Madden, Duggan, Brown and Dabbish 2013). A 2015 study from the same centre shows that only 9% of US adults felt they have ‘a lot’ of control over what information is collected about them and how it is used. This is despite 93% stating they believed that being in control of who can get information about them is important (Madden and Rainie 2015).

Although the widespread adoption of privacy policies represent an attempt to reduce this information gap, research demonstrates that data subjects rarely read privacy policies and in the rare instances when they do, they struggle to understand them. Data protection regulators in the EU and US agree this is an issue and have called for greater ‘transparency’ and ‘control’ to negate the information asymmetry imbalance between data subject and data controller. A number of research studies have used experimental research to gather to assess the impact privacy enhancing features have on reducing information transparency (i.e. reduce information asymmetry). The following section outlines some of the key findings.

2.6.7 Enhancing User Control & The ‘Privacy Paradox’

Research from John, Acquisti and Loewenstein (2010) looking specifically at what triggers a data subject’s information disclosure suggests peoples’ willingness to divulge personal information is influenced by contextual clues and not by rational choice. Findings show different levels of information disclosure in situations where the same potential dangers exist thus demonstrating,

“consumer vulnerabilities in navigating increasingly complex privacy issues introduced by new information technologies” (John et al. 2010, p.858).

Their findings show how contextual cues associated with an increased risk of personal information disclosure can act to suppress privacy concerns resulting in the perverse effect of increasing information disclosure.

“These findings provide support for the idea that cues affect divulgence by rousing, or downplaying, privacy concerns”. (John et al. 2010, p.859).

Related to a phenomenon known as homeostasis or the Peltzman effect, privacy researchers have identified behaviour they have labelled ‘the control paradox’. It is an idea analogous to the user of seatbelts in cars. Although seatbelts do save lives, research suggests that people who wear them tend to drive more recklessly, resulting in increased fatalities amongst cyclists and pedestrians (Peltzman 1975). We see this effect in the context of online privacy

where people may respond to measures intended to protect their personal privacy in ways that counteract the intention of that protection. Brandimarte et al, (2012) note that,

“People often display a notable inconsistency in concern for privacy. They eagerly disclose intimate facts to others, but are bothered when the same information is circulated without their consent. Many online social networks users willfully reveal even self-incriminating information to others – yet, non-substantive changes in privacy policies can generate widespread protests. What can account for these seemingly contradictory behaviors?” (Brandimarte et al. 2012, p.3).

In a series of behavioural experiments, they specifically examine how different levels of perceived control impacts on personal information disclosure and the findings demonstrate sometimes counter-intuitive findings. The authors describe their findings as,

“Consistent with the Peltzman effect, however, we document a “control paradox” such that people who experience more perceived control over limited aspects of privacy sometimes respond by revealing more information, to the point where they end up more vulnerable as a result of the measures ostensibly meant to protect them” (Brandimarte et al. 2012, p.1).

These results have important implications for data protection regulators. They demonstrate cases where features intended to explicitly give data subjects greater control over the release of their personal data, can lead to the unintended consequence of data subjects revealing more personal and riskier information disclosures by data subjects. As such, data protection regulations that are designed to give greater control to data subjects may not have the desired effect on user behaviours that legislators intend i.e. more control can actually lead to less privacy.

An experimental study by Carolan & Castillo-Mayen (2014) examined the 2002 EU e-Privacy Directive rule with regards to data controllers obtaining explicit consent from website visitors (data subjects) to allow them to use cookies to track their behaviour. Their findings show that participants who were given more

control over their privacy options showed a greater willingness to disclose personal information when compared to groups not offered the same control.

“The results demonstrated that the provision of a specific and visible instruction to pay attention to information about cookies had no significant impact on any of the measures assessed. Participants who were presented with this instruction prior to visiting the website demonstrated no greater knowledge of cookies than those who were not. Nor did they show any evidence that they were less likely to trust the website or to disclose information during their visit” (Carolan and Castillo-Mayen 2014, p.377).

These results again demonstrate a somewhat counter-intuitive conclusion i.e. that mechanisms designed by legislators to empower data subjects with greater control may actually encourage disclosure counter to their intentions. The authors believe their findings,

“casts doubt on the efficacy of any legal strategy that assumes that providing visible or salient information about cookies will influence user behaviour, whether positively or negatively, consciously or otherwise” (Carolan and Castillo-Mayen 2014, p.377).

Their research has further implications for companies who have generally resisted EU cookie regulation interventions on the grounds that they are onerous and likely to stifle innovation. The results in fact suggest that facilitating active user engagement with a website may have longer-term benefits in with respect of user willingness to disclose information to them.

Experimental research from Norberg, Horne and Horne (2007) and Taddicken (2014) identify the existence of what they call a ‘privacy paradox’. Observed online behaviours show that levels of actual personal information disclosure across a range of categories (i.e. including financial, demographic and personally identifying) significantly exceeded participants stated intentions to disclose.

“It appears that, in the realm of privacy, behavioral intentions may not be an accurate predictor of actual behavior” (Norberg et al. 2007, p.118).

2.6.8 Privacy Seal Empirics

The Information Commissioners Office (IOC) in the UK describes a privacy seal as,

“a ‘stamp of approval’ which demonstrates good privacy practice and high data protection compliance standards.” (Falmer 2015).

The provision in the GDPR for establishing privacy seal accreditation schemes demonstrate that regulators believe they have a role to play in reducing the information asymmetry that exists between data subjects and data controllers. There is only a small body of empirical research that specifically examines the impact of privacy seals on personal information disclosure and/or purchasing decisions and these studies are outlined below.

In an offline study, Miyazaki and Krishnamurthy (2002) found privacy seals were interpreted by some as signaling a firm’s intention to abide by their stated privacy policies and not violate a data subjects privacy. However, the findings highlight an issue where the mere presence of a privacy seal may,

“create a false sense of trust with no basis in the actual specifics of a privacy policy” (Rifon et al. 2005 p.343).

Their research conducted a series of online experiments to examine the impact privacy seals had on a group of consumers with known higher privacy concerns. The experiment design involved creating a website offering music for sale. Two versions of the website were created, one containing the TRUSTe and BBBOnline privacy seals (described earlier) and one with no privacy seals. A pretest survey contained a scale measuring concern for online privacy and a post-test survey measured participants concern for privacy while they were actually participating in the experiment. The results showed that the presence of a privacy seal increased beliefs in the seal assurances in relation to its information practices (i.e. how data was collected, shared etc). Privacy seals were found to have,

“engendered more trust but did not significantly and consistently influence disclosures of personal information for access to site benefits” (Rifon et al. 2005, p.358).

The findings also showed that privacy self-efficacy moderated data subjects responses to the presence of seals i.e. the more confident participants were in their own ability to protect their privacy the more faith they had in data controller, irrespective of the presence of a privacy seal. Overall the study finds that data subjects,

“may use privacy seals as an efficient assurance that privacy will not be violated. Privacy seals seem to act as a safety heuristic” (Rifon et al. 2005, p.360).

The findings further suggest that,

“Consumers may use privacy seals heuristically to confirm the Web site as “privacy safe” (Rifon et al. 2005, p.340).

However, the authors also warn that,

“Seal programs can only protect privacy if and only if consumers recognise the assurances seal programs afford” (Rifon et al. 2005, p.342).

That is to say, the mere presence of a privacy seal does not mean that a data subject’s personal information will not be shared with third parties. It means that the data controller will only use data in the way the program/regulation that the seal relates to states it is committed to. Two of the leading seal programs, TRUSTe and BBBOnline rely on voluntary compliance although they do engage in ongoing monitoring and offer customer complaint resolution procedures.

Mai, Menon and Sarkar (2010) conducted research into how the use of privacy seals by online vendors impacting on consumers purchasing behaviour. The authors differentiate between what they call ‘scrupulous’ and ‘unscrupulous’

vendors. Scrupulous vendors are those implement and monitor stringent privacy policies and incur additional costs because of this. Unscrupulous vendors do not incur these costs. The research adopts the premise that when making purchases online, consumers face a choice between buying from privacy seal bearing vendors or from non-privacy seal bearing vendors. The latter is associated with an enhanced risk of privacy violation for the purchaser (data subject). There is almost no way for potential consumers to distinguish between the 'scrupulous' and 'unscrupulous' vendors and therefore privacy seals potentially have a role here.

The wider risks to the new media industry from unscrupulous vendors are very real. In Akerlof's (1978) paper examining the second hand car industry, he warns that there of the potential risk of market failure from,

"the presence of people who wish to pawn bad wares as good wares tends to drive out legitimate business" (Akerlof 1978, p.495).

There are costs related to dishonesty and dishonest dealings tend to drive honest dealings out of the market. This should serve as a warning to new media service providers, where unscrupulous (i.e. dishonest) vendors pose a risk to the whole new media market (Doyle 2018). Akerlof believes, "The purchaser's problem, of course, is to identify quality" (Akerlof 1978, p.495). In new media markets, privacy seals/icons have the potential to assist data subjects/service users to identify 'quality' who have been verified to comply with the relevant regulations. Couldry and Turrow (2014) see audience trust as a potential barrier to the seemingly unstoppable momentum of the new media industry.

A Mai et al (2010) study specifically investigates whether privacy seal bearing vendors can charge a premium for the products they sell compared to non-privacy seal bearing vendors. Their study selected a number of 'highly homogenous' products (e-books, audiobooks and textbooks) and found that seal bearing vendors charged a price premium of approx 1.5% compared with non-privacy seal bearing vendors. The findings further suggest that privacy seal bearing vendors are able to charge a premium and that this,

“reinforces the notion that seal programs can provide a viable market-based mechanism for providing privacy protection” (Mai et al. 2010, p.207)

However, the authors note that a major limitation of their research is small number of vendors who carry the privacy seal.

2.6.9 Privacy Icon Empirics

Privacy icons operate in a similar way to privacy seals in so far as they represent a visual indicator of how the data controller may use a data subject's information. Research from Tsai et al (2011) examined how prominent displays of privacy icons impacted consumers' online purchasing decisions. Their research uses a microeconomic framework to represent the consumers' utility maximisation problem. This innovative experimental study was specifically designed to examine the impact of making privacy policy information available to consumers (data subjects) in different way to standard privacy policies. The experiment included the use of a modified search engine that displayed an image/icon indicating the privacy rating associated with each of the merchants listed in the search results. The study found that participants did take the privacy information displayed (via the icon) into consideration when making purchase decisions and were more likely to purchase items from websites with medium or high levels of privacy (as indicated through privacy icon).

Their findings provide evidence that privacy information affects online shopping decision-making and that consumers (i.e. data subjects) may be willing to pay a premium for privacy. They also demonstrate how a,

“business may use their technological means to showcase their privacy-friendly policies and thereby gain competitive advantage” (Tsai, Egelman, Cranor, and Acquisti 2011, p.266)

Firms can strategically manage their privacy settings in ways that fulfil best practice guidelines and maximise profits.

“Our results indicate that, contrary to the common view that consumers are unlikely to pay for privacy, consumers may be willing to pay a premium for privacy” (Tsai et al. 2011, pg.266).

The advertising industry itself is in favour of the use of icons. In 2011, the UK Internet Advertising Bureau (IAB) called for transparency in relation to how and when consumer data was being collected. The IAB called for greater use of

privacy icons to increase trust-building perceptions with users about data collection practices.

“Such tools can offset negative reactions to personalized advertising, enhance website credibility, and make personalized advertising more acceptable” (Aguirre, Mahr, Grewal, de Ruyter and Wetzels 2015, p.44).

However, we also know that marketers have made it clear that they require more and more data from media platforms about audience members (i.e. data subjects) if publishers want to continue to earn revenues in online exchanges (Couldry and Turrow 2014).

2.6.10 *Icons and Seals: Summary*

The empirical research findings discussed have important implications for media firms as they suggest that privacy seals potentially have a role to play in communicating compliance with data protection regulations. The third-party verified seals set out in the GDPR could potentially be standardised to operate in a similar way to the successful energy-labelling scheme. Accredited privacy seals offer increased transparency for media consumers and help reduce the information symmetry that privacy policies alone are failing to address. The GDPR provision for privacy seal accreditation schemes represents an opportunity for data collectors to have their compliance with EU data protection rules externally verified. Evidence from the Tsai et al (2011) study suggests this could also be a source of competitive advantage firms, with consumers willing to pay a premium for such privacy.

However, as we have seen, the research findings in this area are inconsistent. The Leslie, Acquisti and Loewenstein (2010) and Brandimarte, Acquisti and Loewenstein (2012) studies suggest that data subjects can behave in unpredictable and counter-intuitive ways in relation to privacy, i.e. the so called ‘privacy paradox’. The use of privacy seals as envisaged in the GDPR may result in a Peltzman effect where data subjects effectively behave more

recklessly and disclose more personal information. We have also seen that explicit assurances of anonymity can, contrary to their intended purpose, trigger data subjects privacy concerns. This can result in less information being shared in circumstances where the risk of personal information disclosure is lower (Acquisti and Grossklags 2012).

Perceived control also impacts on personal information sharing as it highlights how,

“technologies that make individuals feel more in control over the release of personal information may carry the unintended consequence of eliciting riskier disclosures” (Brandimarte et al. 2013, p.15).

Overall, this research suggests that the more control you give data subjects the more personal information they disclose.

However, it is important to note that in the existing privacy research examined above, the displaying of the seals were used to indicate a firm’s compliance with its own stated privacy policy only. This has led to a miss-interpretation of what the presence of the privacy seal actually represents. For new media consumers, the only way for them to be clear about what would happen to their data was to read the privacy policy for each data-controller (i.e. new media service provider) they visit. This is because data-controllers could all potentially have different privacy policies. We know from literature that this is not practical and that there is evidence to confirm that the vast majority of data subjects simply do not read privacy policies.

The research carried out in this thesis examines the use of privacy seals as envisaged in the GDPR approach. This means that rather than simply acting as a symbol to denote compliance with the data controllers stated privacy policy, the display of an accredited privacy seal represents the validation by an independent third-party that the data-controllers privacy practices are in compliance with the relevant data protection regulation. This provides a benchmark for data-subjects and allows them to differentiate between data-collectors/vendors etc in a meaningful way. Crucially, it also means they do not

need to read each data-controllers privacy policy to understand how their data will be collected and shared etc.

While in the process of designing the experiments conducted in this thesis I sought and received permission to use two 'privacy seals' that are commercially in use. The company ePrivacy operates one such privacy seal accreditation scheme in the EU. Anyone who wants to display their seal must undertake an in-depth audit of their products. They award their privacy seal as testament to,

“a product's compliance with the list of ePrivacyseal EU criteria, which reflects the requirements imposed by EU data protection legislation. This includes the principles of the new EU General Data Protection Regulation.” (ePrivacy 2018)

The company EuroPriSe operates a similar accreditation scheme. Their website states,

“The EuroPriSe website privacy certification is awarded to websites that are compliant with EU data protection law and that meet all of EuroPriSe's high-quality data protection requirements.” (EuroPriSe 2017).

There is no published empirical data on how the display of these third-party verified privacy seals impact on personal information disclosure. If the presence of privacy seals becomes a clear source of differentiation between media firms, this may well precipitate a widespread adoption of privacy seals. However, should the use of privacy seals result in data subjects 'clamming up' and not disclosing personal information, then this has the potential to impact on core revenue streams. The research conducted for this thesis provides insights on how service users behave when a privacy seal is displayed.

2.6.11 *Media Economics: Summary:*

Early approaches to privacy economics in the late 1970's viewed consumers as rational agents who made decisions on their privacy based on a trade (or transactional) approach. Acting with perfect information, users traded their personal information in return for the 'utility' gained from the product or service they were consuming. For example, when applied to a new media service like Facebook, people who set up an account and use the service are in fact trading their personal information in return for services and features that allow them to connect with friends and share information on likes etc. Both sides of the trade are rational actors and making their decisions with full knowledge. The granting of privacy rights through regulation is considered to distort the free market. Any costs associated with adhering to regulation leads to inefficiencies (i.e. inflated prices) and ultimately lower economic growth on a macro level.

From the 1990's, privacy economics literature incorporated more complex concepts like bounded rationality, where framing and heuristics also influence decision-making transactions. Rapid technological advances are seen to have created information asymmetry, where individuals could not understand/consider all potential future uses/sharing of their personal data when making personal information disclosure decisions. To this extent they are not in possession of full information when making decisions. This approach represented a major departure from the existing rational choice theory.

From approx. 2000, the emergence of the Internet and related technologies saw the rapid growth of new media firms. Many of these firms provided the majority if not all of their services for 'free' relying on advertising income for revenue. Data protection regulators recognise the complex trade-offs consumers face when making information disclose decisions and believe that data subjects (i.e. consumers) must be given greater control and transparency of their personal information in order to reduce information asymmetry.

A relatively small but developing body of literature has attempted to gain a deeper understanding of how data subjects make these privacy trade-off decisions. Empirical research outputs demonstrate at times counter-intuitive

findings. Although users express a desire for privacy, their behaviours contradict this i.e. the 'privacy paradox'. Furthermore, privacy-enhancing features that are designed to improve transparency (i.e. reduce information asymmetry) can result in data subjects revealing more personal information in situations where they are more at risk of unwanted disclosure. Further research is needed to improve models of user behaviour and to ascertain the most appropriate policy mechanisms by which consumer privacy might be best served.

2.7 Chapter Summary and Research Gap:

As discussed, academic definitions of privacy drawn from a range of disciplines all see control as *the* key characteristic required to enable an individual to maintain their desired level of privacy. It is important to note that privacy is not the opposite of sharing, it is the *control* over sharing. There may be costs to both parties should too much information be disclosed (Acquisti et al 2016). Data protection authorities agree and see transparency as a key enabler to allow data subjects to exercise control over how and when their personal information is obtained and shared, as well as potential future uses. This is especially relevant in the Internet era where technology advances have reduced barriers to entry and costs, allowing vast amounts of personal data across a range of devices to be collected and stored indefinitely.

Technological development has had a major impact on how media firms operate. The rapid growth of new media service providers has seen the expansion of online services where the provider does not explicitly charge for 'first' content market. This leaves firms completely reliant on the 'secondary' audience markets as a revenue stream and we see new media service providers like Alphabet and Facebook dominate the online advertising market at a time when traditional advertising in sectors like newspapers is in terminal decline. There are benefits to both data subjects and data controllers when certain data is shared. We only have to look to the growth in new media platforms like Facebook and Twitter to see how popular these platforms are. Many of the new media firms who operate these social media platforms are

highly profitable. EU and US data protection regulators agree that the existing data protection provisions in their respective jurisdictions are not adequate. Media economics literature recognises that regulation of the market may be required to prohibit certain types of behaviour within a sector. Regulators believe that lack of adequate data protection poses a risk to consumers (data subjects), service providers (data collectors) and ultimately the wider macro economy. They warn of risks to the growth and sustainability of this industry sector should appropriate provisions to protect personal privacy not exist. EU regulators have attempted to address some of these concerns in the form of the GDPR, which came into force in 2018.

A lack of data protection can lead to economic benefits for data controllers through efficiency gains and increased revenues. This can also bring benefits for data subjects, as firms can pass on efficiency gains through lower prices for products/services. However, this same lack of data protection can also be costly for data controllers through costs and reputational damage associated with data breaches or misuse (Acquisti et al. 2016). Data subjects may be negatively impacted by a lack of data protection through tangible costs such as identity theft and price discrimination as well as by less tangible risks such as psychological discomfort and/or anxiety (Stone and Stone 1990; Feri, Giannetti and Jentzsch 2015).

Since 2000, challenges to the classical rational choice theory argue that privacy decision-making is complex and influenced by other factors like framing and heuristics. Studies show inconsistencies and contradictions in individuals' behaviour when choosing to protect or share personal information. These highlight a dichotomy between attitudes to privacy and actual behaviour, the privacy paradox. They demonstrate that individuals' privacy decisions are complex and differ substantially from each other depending on opinions, attitudes and behaviours. Consumers often lack the required information to make privacy-sensitive decisions and are 'likely' to exchange long-term privacy for short-term benefits (Acquisti, 2004; Acquisti and Grossklags 2005; Acquisti et al. 2016).

Despite contrasting philosophical approaches in terms of regulatory frameworks, European and American regulators have all explicitly stated that current data protection regulations are not sufficient to deal with advances in technology. They identify specific areas that require changes including suggestions for how these changes may operate. In the EU, the new GDPR represents a comprehensive revision of data protection regulations. It explicitly calls for the establishment of a privacy seal accreditation system. No empirical research exists that examines the use of privacy seals that indicate compliance with a specific benchmark regulation. The limited empirical research that looks at the use of privacy seals and icons is outdated and was conducted on US based respondents only.

Research from Evens and Van Damme (2016) shows that while consumers were willing to share basic demographic data like name, date-of-birth, job status and e-mail address, they were more reluctant to share income, contact and financial details. The research conducted for this thesis will explore users willingness to disclose more 'sensitive' data categories (as defined in the GDPR) when a privacy seal is displayed.

As discussed on this chapter, there are a number of gaps in the existing literature and this research addresses some of these gaps. Specifically, it sets out to empirically test the impact privacy seals have on personal information disclosure.

3 Chapter 3: Methodology

This research adopts a mixed methods approach, gathering primary data from surveys and a randomised experimental study. Experiment and survey participants are recruited through the online labour market Amazon Mechanical Turk (MTurk). The survey conducted in relation to RO1 examines whether EU based respondents can identify what categories of data are classified as sensitive personal data under EU data protection provisions. The survey conducted in relation to RO2 asks EU and US respondents to identify the categories of data they believe should be classified as 'sensitive personal data'. The findings from the RO2 surveys are then used in experiments conducted in relation to RO3, to examine personal information disclosure under different treatments.

Experiment participants are randomly assigned a control treatment group(s) where interface changes/manipulations are made. The Fisher Exact Test statistical technique is used to analyse the results. These statistical methods, combined with the experimental design allows for causality claims to be made in relation to the treatment. Specifically, this study examines how the presence of a privacy seal impacts on personal information disclosure. The following sections articulate how the research objectives of this thesis have shaped the philosophical, methodological, method and data analysis approaches adopted.

We will begin by briefly recapping on the research question and research objectives of this study.

The overarching research question this thesis aims to address is the following:

What impact(s) do changes to data protection regulations have on personal information disclosure in online environments?

In order to answer this research question, the following research objectives are addressed:

RO1 – To examine if respondents can identify the types of data categorised as ‘sensitive data’ status under current EU data protection regulations.

MTurk is used to recruit survey respondents based in EU member states only. This is because only EU data protection explicitly declares some types of data as sensitive personal data. Categorising data as sensitive imposes restrictions on how that data is collected and processed i.e. what it can be used for. Sensitive data includes any data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership as well as data concerning health or sexual orientation.

EU data regulators see sensitive data as an important issue for data subjects as,

“misuse of these data could have more severe consequences on the individual’s fundamental rights, such as the right to privacy and non-discrimination, than misuse of other, “normal” personal data” (Article 29 Working Party 2011, p.4).

The US regulation authorities take a similar view stating that,

“disclosure of sensitive information can violate individual rights, cause injury or discrimination based on sensitive personal attributes, lead to actions and decisions taken in response to misleading or inaccurate information, and contribute to costly and potentially life-disrupting identity theft” (White House 2012, p.100-101).

The survey results will provide a useful overview of levels of awareness in relation personal data categorised as ‘sensitive personal data’ under current EU data protection regulations. The results may also provide insights from a public policy perspective. Do regulation authorities need to do more to inform its citizens about why these categories are classified as sensitive personal data? Are there categories of data that people believe are sensitive (and therefore afforded more protection) when they are not? The results also provide further

context in relation to RO2 where EU and US respondents are asked to what categories of data should be categorised as sensitive data.

RO2 – To examine what categories of personal data respondents consider ‘sensitive personal data’.

While RO1 examines respondents’ knowledge of data that is currently classified as sensitive data under EU data protection regulations, RO2 sets out to identify what categories of data respondents *believe* should be classified as sensitive data.

A number of White papers and reports that preceded the publication of the GDPR proposed expanding the list of categories of personal data that are categorised as ‘sensitive’ data. However, in the end the GDPR only expanded the list to include genetic and biometric data.

MTurk service is used to recruit survey respondents from EU member states and the US. As discussed, sensitive data categories were initially set out in the 1950 ECHR and these were subsequently adopted into relevant EU data protection regulations.

The purpose of conducting this survey is to examine what categories of data EU and US respondents consider ‘sensitive’ personal data is threefold.

Firstly, the global nature new media firms means it is useful to compare the types of data respondents think should be sensitive and compare common grounds and differences across EU and US jurisdictions.

Secondly, if a significant number of respondents believe that certain categories of personal information that are not sensitive data should be considered as such, then will may help to inform future public policy debates.

Thirdly, and most significantly in the context of this thesis, the categories of data that respondents identify as most sensitive personal data will then be used in the experiments conducted in this thesis (see RO3 described below).

RO3 – Conduct experiments to test if use of ‘privacy seals’ effect data subjects personal information disclosure.

Although the GDPR calls for the establishment of privacy seal accreditation schemes, there is no EU based empirical research and only limited US based empirical research relating to the impact privacy seals have on personal information disclosure. Widespread adoption of the use of privacy seals by data collectors is unlikely in the absence of an evidence-based understanding of how privacy seals impact on personal information disclosure.

Using the results of the survey conducted in relation to RO2, MTurk workers are asked to participate in the survey on business ethics. At the end of the survey, respondents are asked to complete some questions about to, ‘tell us a bit about yourself’ (i.e. their age, gender etc). They are also asked questions that will reveal details identified as ‘sensitive’ data in the survey outlined in RO2 above. Respondents are assigned to control and treatment groups where the treatment applied represents very minor changes to the survey presentation. Using appropriate statistical techniques we can establish whether a causal relationship exists between the treatments and the personal information disclosed by the experiment participant.

This experiment was conducted separately on both EU and US audiences. Again, the global nature of new media firms means it will provide a useful comparison of results between the different jurisdictions.

3.1 Research Philosophy

This research adopts a positivist research philosophy. The positivist approach can be described as a,

“position that advocates the application of methods to the study of social reality and beyond” (Bryman and Bell 2003, p.14).

Knowledge is derived by the gathering of facts, which, through the process of deduction provide the foundations for the creation of new knowledge. The Positivist approach relies on a quantitative approach, collecting data to be tested through statistical techniques and is free from emotional or feelings in their interpretation. The results further inform the underlying theory which may in turn be tested through further research.

3.1.1 Ontology

From an ontology perspective, this research adopts a behavioral economics approach with respect to human behaviour as it relates to media consumption. Individual choice and decision-making are viewed as not stable and may be subject to influence from rules, emotions, framing, desires, and other things (Angner 2015). This approach contrasts with, and rejects, the neoclassical reductionist view where individuals are considered to behave in a consistent manner (i.e. profit-maximizing) with respect to their preferences and decision-making. As we saw in the literature review, early economic approaches to privacy assumed individuals behaved in this reductionist approach i.e. behaved as rational agents. However, privacy economics literature developed over the years to take on themes from behavioural economics, recognizing the complexity individuals face when in their decision-making.

Behavioural economics recognizes that individuals face cognitive limits when making decisions (Sharif and Mullainathan 2013). Faced with uncertainty, a lack of full information means individuals' decision-making is based on heuristics, i.e. they behave in a utility-satisfying way rather than utility-maximising way. Whether individual preferences are stable and rooted inside individuals or

whether they are influenced by external factors is still a contested question (Henrich, Boyd, Bowles, Camerer, Fehr, Gintis, and McElreath 2001).

As data protection regulators have identified, the speed of technological development in recent years (Oliver 2014; Oliver 2018; Oliver 2018b; Küng 2018) means individuals face complex cognitive challenges when using these technology services. They also recognise that privacy policies are not effective in communicating what data is being collected, for what purpose and who it will be shared with etc. This lack of transparency and control is something the GDPR sets out to address. The establishing of a privacy seal accreditation scheme is something the EU regulators have called for to address the lack of transparency and control.

A number of privately held (i.e. not part of a formal EU institution) European based businesses operate privacy seal accreditation schemes. In broad terms, the scheme works as follows. Any firm can apply for a privacy seal from one of a firm that offers an accreditation scheme. The accreditation firm then works with the applicant and checks all relevant data protection provisions relating to data collection, storage and processing systems to ensure these practices comply with the relevant data protection regulation in that area i.e. the GDPR. Firms that demonstrate compliance with the GDPR are 'awarded' the privacy seal and can display the seal demonstrating third party verified compliance.

There is an up-front fee related to this assessment process. If the applicant is successful, they receive permission to use the accredited privacy seal. There is generally an annual fee charged by the accreditation firm and with monitoring ever 6, 12 or 18 months to ensure the holders data practices remain within the relevant regulatory requirements.

The core goal of this research is to examine if the presence of privacy seals have an impact on personal information disclosure. Related research in this area suggests that the presence of privacy enhancing icons/symbols triggers privacy concerns and results in at times somewhat counter-intuitive effect of reducing personal information disclosure. Therefore, from an ontology perspective, behavioural economics is a valid way to examine how decision-

making in relation to personal information disclosure is affected (or otherwise) by the presence of a privacy seal.

3.1.2 Epistemology

Epistemology is the examination of, “how we make knowledge” (Dillon and Wals 2006, p.550). Rejecting the neoclassical view that individuals exercise consistent rational choice with regards to preferences, behavioural economics instead sets out to test theories with regards to decision making by means of experiments. This approach reflects the development of privacy economics literature discussed in the literature review where early rational choice models are superseded by approaches that account for bounded rationality and recognise the complexity in decision-making.

Nobel Prize winning economist Daniel Kahneman explains how a behavioural economics approach explores,

“systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models” (Kahneman 2003, p.1449).

He describes different two different ‘systems’ that are involved in decision-making, the ‘intuitive’ and the ‘reasoning’. The intuitive system is characterised as fast, effortless and often inconsistent with regards to choices made. The reasoning system is more elaborated, slower and more reliable in terms of the consistency of choices. The intuitive system deviates from the type of behaviour a rational behavioural model would predict, we assume that individuals are biased by bounded rationality. This epistemological realism assumes human behaviour can be observed and described by scientists. With regards to classifying empirical findings, the assumed behaviour of individuals is seen as the benchmark for analysing observable behaviour. We can therefore measuring observable behaviour and analyse how far these deviate from the benchmark.

3.2 Research Methodology

This is an interdisciplinary study incorporating philosophies and norms from the media economics and law disciplines. An interdisciplinary approach is appropriate as the research objectives involve,

“answering a question, solving a problem, or addressing a topic that is too broad or complex to be dealt with adequately by a single discipline or profession... [It] draws on disciplinary perspectives and integrates their insights through construction of a more comprehensive perspective” (Klein and Newell 1998, p.393-394).

The data collection for this research was carried out in 2016 and 2017.

From a methodological perspective, this thesis employs a quantitative methodology. Measurement is a key emphasis of a quantitative approach. The use of quantifiable variables combined with appropriate statistical method gives researchers the ability to identify more precise estimates of the differences between respondents and/or groups of respondents. Quantitative research goes beyond description of how things are and is concerned with causality i.e. why things are the way they are (Bryman 2012).

It is important to note that while a quantitative approach is often associated with the development of hypothesis, this is not always the case as,

“a great deal of quantitative research does not entail the specification of a hypothesis” (Bryman 2012, p.161).

Rather than develop specific predictions in the form of hypothesis, this research applies appropriate statistical methods to interpret the empirical evidence gathered from the survey and experiments.

Consistent with both our epistemological realism and ontological approach, a quantitative approach enables the researcher to potentially make causal claims in relation to the independent and dependent variables used in experiments.

However, quantitative research is not without its critics. Phenomenologist Schutz (1962) believes that while a quantitative approach might be appropriate for the natural world, it is not useful when applied to the social. The core accusation is that the quantitative researcher approaches fail to distinguish individuals and social institutions from the natural world.

Cicourel (1964) argues from a similar perspective and points out how respondents may interpret the meaning of key phrases in questionnaires differently. This may result in the concepts social scientists attempt to measure being assumed and therefore not real. Suggestions that the inclusion of fixed-choice answers mitigates this risk merely demonstrate,

“a solution to the problem of meaning by simply ignoring it” (Cicourel 1964, p.108).

He questions whether respondents have the required knowledge to answer the question being addressed to them.

The GDPR calls specifically for the establishment of a privacy seal accreditation scheme. However, it does not cite any empirical evidence to suggest whether such a system will in fact deliver the improved information transparency and control that regulators believe is missing from current data protection regulations. This research sets out to address the research gap that exists in this area. It is important to note that all methodological approaches have their drawbacks and as discussed above. Similarly, quantitative approaches have both its critics and supporters. As we will see later in this chapter, the use of randomised treatment and control groups is used to help isolate the effect the presence of a privacy seal has on personal information disclosure.

3.3 Research Methods

“A research method is simply a technique for collecting data” (Bryman 2012, p.46).

Consistent with a positivist research philosophy, this research employs online randomised experiments to examine how features proposed by data protection regulators (i.e. privacy seals) to improve data subjects control and transparency over their personal information impacts on the actual behaviour of data subjects. A number of surveys are also conducted examining if respondents can identify what types of data should be categorised as sensitive personal data. Categories of data that the highest percentages of respondents consider ‘sensitive personal data’ are then used to inform specific elements of the experiments.

The experiments conducted here help to redress this imbalance in data collection methods. They examine individuals’ personal information disclosure and specifically to test if the display of privacy seals as ‘kite-marks’ impacts on personal information disclosure. Participants are randomly assigned to control and treatment groups. The control group represents the benchmark for our analysis. We then apply specific features or ‘treatments’ to examine how they impact on personal information disclosure. .

From a methodological perspective, this research gathers data from two sources. They are:

1. Surveys: In total, three online surveys are carried out to answer RO1 and RO2. The findings establish whether ‘sensitive data’ is a concept respondents are familiar with i.e. which categories of data are considered ‘sensitive’ under existing EU regulations. The results also inform us as to which categories of personal data EU and US based respondents believe should be classified as ‘sensitive personal data’. The results of the surveys feed into the design of the experiments described below.

2. Experiments: Two online experiments are conducted in relation to RO3. They test if displaying a privacy seal impacts on personal information disclosure. The experiments take the form of an online survey but crucially they also include control and treatment groups. The use of treatment and control groups gives the potential for causal claims to be made in relation to independent and dependent variables.

The following sections provide justification as well as explanation of the structure and composition of each method.

3.3.1 Online Survey Method

New media firms have played a significant role in the transition of many developed countries from an industrial society to an information society.

“Our “society,” thus, requires a prompt and accurate flow of information on preferences, needs, and behavior. It is in response to this critical need for information on the part of the government, business, and social institutions that so much reliance is placed on surveys” (Scheuren 2004, p.7).

This research conducts three surveys directly related to RO1 and RO2.

“Low-cost computing and rapid development of technology have created new environments for conducting survey research” (Sue and Ritter 2012, p.1).

This research uses one such *new environment for conducting research* in the form of the ‘crowd-sourcing’ application MTurk to recruit survey (and experiment) participants. Research from Dupagne (2018) shows that over the period 2004-2016, many media management researchers used Internet based samples, often due to their cost effectiveness. However, as with many research methods, there are advantages and disadvantages in relation to the use of online surveys.

Online surveys offer a number of advantages as a research method. First and foremost they can be very quick to administer. For example, for one of surveys carried out for this research, the MTurk platform was used to recruit hundreds of survey participants in less than 48 hours. Online surveys also have much lower constraints in terms of geographical coverage and as this thesis examines the differences between EU and US citizens this was an important consideration in terms of participant recruitment. Sue and Ritter (2012) highlight another potential advantage of online surveys. As no researcher is present, respondents can answer at their own pace, resulting in more honest answers to questions. These were all important considerations in this research.

Critics of online surveys highlight the fact that as participants are not randomly selected, we cannot generalise findings. For example, as not everyone has Internet access, some socio-economic groups are excluded from the sample data collected. Furthermore, as the number of people who actually have Internet access is both difficult to estimate and changing it is difficult to state how representative this type of sample is. This issue is somewhat mitigated by the fact that the research we are conducting is specifically related to online behaviour so the fact that participants require Internet access is not a significant issue. Research from Clifford, Jewell and Waggoner (2015) showed MTurk ‘workers’,

“tend to be more politically liberal, younger, less religious and less racially diverse” (Clifford, Jewell and Waggoner 2016, p.1)

as compared to the general US population. Clearly this would be more of an issue if one is conducting politics related research, but it is also useful to be aware of generally.

Participants in this research were recruited from MTurk and therefore represent a non-probability convenience sample (Dupagne, 2018). In later sections of this chapter we will address this issue of generalisation in more detail. However, this issue is somewhat mitigated by the fact that the focus of this research is to

examine how the presence of privacy seals impact on online personal information disclosure so limiting the sample to Internet users is justified.

A more detailed description of the steps including screen shots of how the survey was presented is detailed in the Appendices. However, the survey operates broadly as follows:

Overview: How the Surveys Operated:

The exact criteria used to select MTurk workers to participate in these surveys are detailed in later sections of this chapter. However, broadly speaking they operated as follows:

- Registered MTurk 'worker' who meets the experiment criteria is presented with a link to a survey with the title '*A survey about data protection*'
- Participant selects the hyperlink to the survey.
- All participants presented with the same 'Overview' page explaining what the survey is about. They select the <Next> button to continue.
- Participants are then presented with a '*Survey Instructions*' page briefly telling them what to do.
- Participants are presented with questions relating to the relevant research objective.
- The experiment then asks respondents for some additional information about themselves including age, gender etc
- Participant selects the <Finish Survey> button and they are presented with a 'completion code' and they enter this on MTurk application to confirm completion of the task.

The results of these surveys inform what categories of personal data requested in the experiments detailed below.

Please see Appendix chapter at the end of this document for links and example screenshots for the surveys and experiments carried out for this thesis.

3.3.2 Experiment Method

Analysis of methodological approaches the media management journals *International Journal of Media Management* between and the *Journal of Media Economics* publications from 2004-2016 by Dupagne (2018) shows that,

“the three most frequent used data collection methods at the aggregate level were secondary data (32%), survey (25%), and literature review (14%). Most of the other research methods (experimental, legal analysis, qualitative content analysis, field observation, focus groups and historical method) were scarce in these two journals from 2004 to 2016. (Dupagne 2018, p.369).

Online experiments are carried out in order to answer RO3. As a research method, one of the primary advantages experiments offer is,

“a simple approach for determining cause and effect, and it has produced convincing findings on a wide range of substantive topics” (Gaines, Kuklinski, and Quirk 2007, p.5).

Varian (2014) argues that experiments represent,

“the gold standard for causality. More specifically, what you want are randomly assigned treatment-control experiments.” (Varian 2014, p.29).

Causality can also be inferred from cross-section and/or panel data. However, these approaches pose a number of methodological challenges including selection biases, mutual causation and correlated measurement errors. The

survey experiment avoids many of the issues associated with more traditional cross-sectional and panel data research approaches (Gaines et al 2007).

An additional advantage of an experimental methodology is that,

“Randomized experiments are more likely to yield unbiased estimates of causal effects than typical observational studies because the randomization of treatment and control groups equal on average in terms of all (observed and unobserved) characteristics” (Horiuchi, Imai and Taniguchi 2007, p.669).

Importantly, “experimenters usually intend their studies to reveal the workings of the real world” (Gaines et al 2007, p.9).

In research that focuses on economic approaches to privacy, experiment methodological approaches are common adopted.

Experiment participants are presented with statements and asked to select an answer from a list of options in a drop down menu. Through the use of treatment and control groups, the experimental design can isolate the impact of a policy measure or change of situation i.e. the presence or otherwise of a privacy seal and or alternative logo/kite-mark. Applying appropriate statistical techniques to the results we can accurately measure if a causal link between policy measures and user behaviour exists. These types of empirical investigations create new and/or extend existing economic models.

Randomised experiments are built on 4 key principles:

- **Control:** Differences between treatment and control groups are kept to a minimum for each experiment so as to control for possible confounding effects. In this research, exactly the same questions are presented to the participants. The only difference between control and treatment groups are the images displayed in the header and footer of the web pages displayed

- **Randomisation:** Participants will be randomly assigned to treatment and control groups. The randomisation to groups is done automatically by the Qualtrics software and the participant does not get to choose their treatment group, is not aware of what group they have been assigned to nor is the participant aware that other groups exist.

- **Replication:** The higher the number of cases observed the more accurately we can estimate the causal effect of the explanatory (independent) variable on the response (dependent) variable. All control and treatment groups have a minimum of 70-80 participants in each.

- **Blocking:** If we know or suspect variables other than the treatment variables will influence the response variable then we can choose to group individuals based on these blocks i.e. blocking. Blocking was not used in the experiments conducted here.

(Diez, Barr and Cetinkaya-Dundel 2012)

Data collection for experiments may encompass a number of steps.

“In experimental research, this is likely to entail pre-testing subjects, manipulating the independent variable for the experimental group, and post-testing respondents” (Bryman 2012, p.162).

The methodology of this thesis follows many of these steps. The RO2 survey described above establishes the categories of data respondents believe should be categorized as sensitive personal data (a sort of pre-test). We then incorporate these findings into the experiments where the independent variable is an image that is then presented/manipulated in the treatment(s).

While experimental research has many advantages and allows us to make causal links, like any method, it also has its limitations. Among what they describe as ‘Practices That Could Produce Misleading Inferences’, Gaines et al

(2007) highlight how 'single-shot survey experiments' fail to measure the duration of their effects. They question whether survey experiment findings are relevant if the treatment effect no longer persists shortly afterwards. The authors distinguished between accepting a message and learning from that message. Drawing on political science literature, they cite examples of how messages seen by experiment participants from 'highly credible sources' had a larger immediate effect than less-credible sources. They identified a so-called 'sleeper effect', where credible sources increased short-term acceptance but did not improve long-term learning (Hovland and Weiss, 1951). A further study in this area from Druckman and Nelson in 2003 found that framing effects in their experiments had dissipated after 10 days.

Gaines et al (2007) identify a further complication that survey experimenters face in relation to the contamination of experimental settings. They argue that in order for experimental research to have merit, the effects that experiments seek to uncover must occur or have occurred in the real world i.e. if the effects had never occurred then they would be no motivation for the research. This means that respondents are likely to have come across the treatment before and therefore have been contaminated by their previous experiences. Experimenters therefore must be aware that respondents are likely to have already participated in a similar real world experiment and that this may impact the treatment effects observed.

While there is no doubt that the limitations discussed above do apply to respondents recruited on MTurk, they also apply to all behavioural experimentation approaches. Paolacci, Chandler and Ipeirotis (2010) looked at tradeoffs across a range of experimental recruitment methods including laboratory, traditional web study and a web study using a purpose built and MTurk. As you can see from the summary table below, across a range of potential recruitment issues like non-response errors, risk of dishonest errors etc, MTurk actually holds up quite well.

Figure 3-1 Trade-off of Different Recruitment Methods

	Laboratory	Traditional web study	Web study with purpose built website	Mechanical Turk
Susceptibility to coverage error	High	Moderate	Moderate	Low
Heterogeneity of samples across labs	Moderate	High	High	Low
Non-response error	Low	High	High	Moderate
Subject Motivation	Moderate / High	Low	Low	Low
Risk of multiple responses by one person	None	Moderate	Moderate	Low
Risk of contaminated subject pool	Moderate	High	Moderate	Low
Risk of dishonest responses	Moderate	Low	Low	Low
Risk of experimenter effects	Low	None	None	None

(Source: Paolacci et al 2010, p414)

Overview: How the Experiments Operated:

The exact criteria used to select MTurk workers to participate in these surveys are detailed in later sections of this chapter. However, broadly speaking they operated as follows:

- Registered MTurk ‘worker’ who meets the set criteria sees a link to a survey with the title ‘*A survey exploring ethical behaviour*’
- Participant selects the hyperlink to the survey.
- All participants presented with the same ‘Welcome Page’ and are asked to select <Next> button to begin.
- Participants see an overview page telling them there are 2 parts to the task.

- The Qualtrics software randomly assigns the participant to one of the treatment/control groups. Note: the participant is not aware that there are other groups or which group they have been assigned to.
- All experiment participants (irrespective of group) are presented with the exactly the same survey questions with the same sets of answers available in the drop down menu of responses available. The only manipulations in the user interface will be the presence of logos and/or privacy seals.
- The experiment then asks respondents for some additional information 'about themselves' including age, gender, education and income etc
- Participant selects the <Finish Survey> button and they are presented with a 'completion code' and they enter this on MTurk application to confirm completion of the task.

In order to determine the causal effects, the only differences between the treatment and control groups are very minor changes (or manipulations) to the user interface. In the experiments conducted for this research the manipulation involves changing the independent variable displayed in the header and/or footer to display a different image (or no image). If the outcome (dependent variable), which in our case is the categories of personal information that participants choose to disclose about themselves, are different between control and treatment groups, we claim that these are a result of the user interface manipulation i.e. our independent variable. The Fisher Exact Test is used to determine if the differences in effects/responses across control and treatment group's are statistically significant.

3.4 Sample

As the research conducted in this thesis examines the impact privacy seals have on personal information disclosure, the potential target population is anyone who enters personal information on any new media service on the Internet.

“Population refers to the entire group of people, events, or things of interest that the research wishes to investigate” (Sekaran 1992, p.225).

Accurately estimating the population frame relevant to the target population for this is very difficult due the geographical spread and dynamic nature of potential relevant citizens (i.e. those with Internet access) across multiple countries. This complexity makes the construction of a valid subset/sample by way of simple random sample extremely difficult.

The absence of a clearly defined target population means we cannot accurately estimate the odds (or probability) of a person participating in our survey. The participants in surveys and experiments in this research are all recruited through the Amazon Mechanical Turk application and this therefore represents a non-probability convenience sample. A non-probability sample does not mean that the sample is not representative of the population but it does mean that we cannot generalise our results to the wider population (Sue and Ritter 2012).

Sample size is also an area to consider. Roscoe (1975) proposes a number basic ‘rules of thumb’ for determining sample sizes. Firstly, sample sizes greater than 30 and less than 500 are appropriate for most research. Secondly, where samples are broken into subsamples (i.e. treatment groups etc), a minimum sample size of 30 in each subcategory is necessary. All of the surveys and treatments in this thesis meet these approximate rule of thumb sample sizes.

3.4.1 Sample Recruitment

“Online labor markets are ideal for conducting incentivised behavioural experiments” (Rand 2012, p.174).

Amazon Mechanical Turk (MTurk) is one such online labour market. It operates as,

“a crowdsourcing web service that coordinates the supply and demand of tasks that require human intelligence to complete” (Paolacci et al 2010, p.410).

Broadly speaking the online MTurk marketplace operates as follows. Potential employers (requesters) post Human Intelligence Tasks (or HITs) in return for set rewards (payment in US\$'s). In this case the HIT is a survey or experiment but other types of HITs may include writing product descriptions or reviews, identifying music performer etc, basically anything that requires human intelligence to complete. MTurk has a database of thousands of registered employees (workers) from all over the world. Requesters post HITs that include a short description, the payment being offered as well as a preview of the task. The MTurk software allows requesters to specify criteria that workers must meet in order to be permitted to complete a HIT. The criteria settings available include gender, approval rating, workers country of residence, how many times they would like the HIT completed etc. Then, when workers access their MTurk account, they only see a list of HITs where their criteria matches the criteria set by the requester.

HITs can vary in length of time but generally take only a few minutes to complete. The requester can check to ensure the task has been completed correctly before authorising payment to be made. Conveniently, the software generates a unique code for each HIT and workers are asked to enter this code on the MTurk application when the task is complete. Requester's pre-load money into their AMT account and the software enables a straightforward worker batch payment authorisation system. Requesters can also choose to pay bonuses to workers who have or alternatively can refuse payments workers who they believe have not completed the task correctly. Requesters who refuse to pay workers for any reason run the risk of being listed on some of the 'workers' community forum websites. This can result in requesters being blacklisted and therefore avoided by sections of workers.

All of the respondents to the surveys and experiments conducted for this thesis were recruited through MTurk. The actual survey questions and control groups for the experiments were all created using the Qualtrics survey application to which Bournemouth University has a multi-user licence agreement. The following sections set out some key overview data including when the surveys/experiments were conducted, how many participants were recruited, and how much they were paid etc.

3.4.1.1 MTurk ‘workers’: Criteria

The MTurk application allows you configure certain criteria that ‘workers’ have to meet in order to participate in your Human Information Task. Only those workers that meet your set criteria are given the option to select the link to that HIT. Setting worker criteria is a useful way to ensure you collect quality data from your respondents. Existing research in this area offers guidance on what settings should be considered and these are set out in the ‘Validation of Data’ section of this chapter (below).

All survey and experiment participants in this research were required to meet two core criteria. These were; (a) ‘HIT Approval Rate’ greater than 95 and; (b) their ‘Number of HITS completed greater than 100’. Additional ‘other’ criteria, such as location and award (payment) did vary depending on the survey/experiment and these are set out below. Initially it was difficult to know what price to pitch the assignments at. A couple of test pilots/trials were run at \$1.00 and then at \$0.50 per assignment and eventually the payment was set at \$0.65 per HIT was decided on.

The tables below provide a summary of the survey and experiment recruitment through MTurk.

Figure 3-2 RO1 Survey: Sensitive Data (EU)

RO1 Survey: Sensitive Data - current categories (EU)	
Total Responses:	153 responses
Title:	A survey about data protection
Description:	What types of personal data are categorised as sensitive data
Reward:	\$0.65
Time allowed:	1 hour
Location:	EU based workers
Started:	July 2016
Ended:	July 2016

N

Note: As this survey relates to current EU categories of sensitive it recruited participants from EU member states only.

Figure 3-3 RO2 Survey 1: Sensitive Data (EU)

RO2 Survey_1: Sensitive Data – new categories (EU)	
Total Responses:	225 responses ⁶
Title:	A survey about data protection
Description:	What types of data do you consider as sensitive data?
Reward:	\$0.65
Time allowed:	1 hour
Location:	EU based workers only
Started:	30 th July 2016
Ended:	31 th July 2016

Figure 3-4 RO2 Survey 2: Sensitive Data (US)

RO2 Survey_2: Sensitive Data – new categories (US)	
Total Responses:	201 responses ⁶
Title:	A survey about data protection
Description:	What types of data do you consider as sensitive data?
Reward:	\$0.65
Time allowed:	1 hour
Location:	US based workers only
Started:	30 th July 2016
Ended:	31 th July 2016

Figure 3-5 RO3 Experiment 1 - Privacy Seal (EU)

RO3 Experiment_1: Privacy Seal Experiment (EU)	
Total Responses:	337 responses ⁶
Title:	A survey exploring ethical behaviour
Description:	Answer questions about personal ethical behaviour.
Reward:	\$0.80
Time allowed:	1 hour
Location:	EU based workers only
Started:	2 nd Feb 2017
Ended:	4 th Feb 2017

Figure 3-6 RO3 Experiment 2 - Privacy Seal (US)

RO3 Experiment_2: Privacy Seal Experiment (US)	
Total Responses:	312 responses ⁶
Title:	A survey exploring ethical behaviour
Description:	Answer questions about personal ethical behaviour.
Reward:	\$0.75
Time allowed:	1 hour
Location:	US based workers only
Started:	21st Feb 2017
Ended:	21stFeb 2017

3.5 Research Design

Reliability, replication and validity are all key criteria we must consider in social research design. The following sections set out how the surveys and experiment elements in this thesis are designed.

3.5.1 Online Survey Design

The best survey questionnaires,

“look professional and motivating, are easy to comprehend, are inviting and not intimidating, and are accessible to everyone in the target population” (Sue and Ritter 2012, p.76).

With this in mind, the survey and experiments in this thesis followed many of the key design principles for online survey design as described by Dillman (2000). This includes giving due consideration to important areas such as welcome screen, access control, first question, conventional format, colour, instructions, formats of response options, navigation guides, text size and font and appearance on multiple devices.

All of the surveys carried out for this research displayed a welcome screen to provide a brief introduction to the survey. Dillman (2002) suggests that when working with probability samples it is necessary to issue a password to ensure unwanted respondents do not get access to the survey. However, as discussed earlier, respondents in these surveys and experiments represent a non-probability sample and as such a password was not required for those who took part in the surveys. That said, it is worth noting that MTurk workers who taking part with a unique code to enter when the survey is completed. The MTurk software application prevents the same worker from participating in a survey more than once and thus duplication is avoided.

Dillman (2002) believes that displaying a short first question in a survey is considered useful as this helps set the tone for the rest of the survey (see section on ‘Survey Questions’ below). If the first question is complicated or

unclear in terms of what the respondent is expected to do, this may lead to assumptions that this is indicative of the rest of the survey, resulting in high survey abandonment rates. Presenting the first and indeed all questions in a conventional format including left justified text and answer options appearing to the right of or directly below the question asked is considered desirable and these conventional formats are included in the survey design.

Colour is also an important consideration and Qualtrics templates were used with a neutral/white background so as to avoid potential issues with colours displaying differently on different types of devices. The Qualtrics application automatically formats the survey so that it displays correctly on mobile devices.

Instructions are also provided for all surveys providing brief directions to respondents. All survey questions are answered via radio buttons (for demographic related questions) or via drop down menus and all questions are closed-ended. Next and back buttons are available on pages and all questions and all survey text is presented in Ariel font size 12.

Navigation buttons are displayed and respondents can navigate using back and next or start buttons as required. As all of the surveys are relatively short a status bar was not included to show the percentage complete etc.

The survey design in this thesis has adopted many of Dillman's (2002) design guidelines, adapting them where appropriate.

3.5.2 Survey: Questions

“A survey question is a measurement tool, a way for researchers to discover a respondent's opinion, knowledge and behaviour” (Sue and Ritter 2012, p.51).

The questions in the surveys (and experiments) in this thesis are closed-ended questions where respondents must select an answer from the range of options presented. There are two types of closed-ended questions. Firstly, there are a number of dichotomous questions (i.e. they only offer two possible responses

such as yes/no, or males/female etc). Secondly, a number of multiple-choice with a range of mutually exclusive answers available to select are used. Closed questions offer a number of advantages including the easy of processing the answers from a coding. This in turn makes the subsequent analysis easier to conduct. Closed questions are easier and quicker for respondents to answer. However, closed questions do by their very nature limit responses and this may result in potentially interesting replies not being recorded. Care also needs to be taken to ensure there is no overlap in the answer options provided (Bryman 2012).

All survey and experiment questions in this research force respondents to select one answer only from a drop down list. Existing research show this is preferable to a 'check all that apply' format in online surveys (Smyth, Dillman, Christian, & Stern, 2006). They found that respondents took longer to answer forced choice questions compared to 'check all that apply' questions suggesting a deeper level of processing in response to the options presented. In the experiments conducted for this research, the demographic questions all include a 'would rather not say' option. This was included in light of the core research objective examining what impact the presence of certain icons would have on information disclosure. So rather than force respondents to select an option in cases where they did not want to disclose the 'would rather not say' option was included.

Findings from Krosnick, Holbrook, Berent, Carson, Hanemann and Kopp (2002) found that offering a 'no opinion' may discourage respondents from sharing their true feelings on a subject and it was therefore not included as an answer to any of the survey questions addressed in this thesis. This author argues that a 'would rather not say' option is different to a 'no opinion answer'.

Our survey requires the participants to answer an 11-question survey where each of the questions asks:

“How do you believe data revealing your [insert category 1-11] should be classified under data protection laws?”

The choice of answers available to select from the drop down menu were:

- Sensitive personal data
- NOT sensitive personal data
- Not sure / do not know

Please see Appendices for detail and screenshots of actual survey.

The results of this survey element are analysed and the four categories of data the highest percentage of survey respondents identify as 'sensitive personal data' are identified. These are the categories of data that the experiments (detailed in next section) will request from respondents.

3.5.2.1 Survey: Demographic Questions

Much of the discussion on survey questions above is also relevant to demographic questions. We consider them separately for two main reasons,

“(a) they are ubiquitous on survey questionnaires, and (b) many people consider these items to be sensitive in nature” (Sue and Ritter 2012, p.69).

Demographic questions typically include questions relating to gender, age and level of education. This data allows for the result to be analysed and examine if a particular gender, age group and/or level of education considers certain categories of data as sensitive particular data.

3.6 *Experimental Research Design*

“Experimental research is frequently held up as a touchstone because it engenders considerable confidence in the robustness and trustworthiness of causal findings. In other words, true experiments tend to be very strong in terms of internal validity” (Bryman 2012, p.50).

Existing experimental studies in this subject area often choose not to present their studies as specifically addressing attitudes to privacy as this is considered to trigger privacy concerns in participants and therefore distort results through non-response errors etc. In order to mitigate against pre-existing privacy beliefs that may result in self-selection biases, experiments are often presented as surveys exploring 'attitudes to business ethics' or similar (John et al, 2010; Tsai et al, 2011; Acquisti and Grossklages, 2012; Brandimarte et al, 2013). This research adopts a similar approach and respondents will not be explicitly told they are participating in a study that has anything to do with privacy or data protection. The following sections will detail the key elements that required in experimental research design.

3.6.1 Experiment Design

A 'classical' experimental design incorporates a number of key elements including manipulation, control and treatment groups' assignment as well as random assignment to those group(s). A true experiment is one where the independent variable is manipulated in order to determine if it has any effect on the dependent variable. In the experiments conducted for this research, the independent variable is an image (i.e. Privacy Seal and or BU Logo) and the dependent variable(s) are the four sensitive personal data fields presented to participants. The experiments conducted here are post-test only design and there is no experiment pre-test.

Experiment participants are randomly allocated to the experimental groups (control or treatment), where the independent treatment variable has been manipulated. As the only difference between the treatment and control groups is the manipulation of the independent variable, we can analyse how the changes to independent variable (i.e. the manipulation of the treatment group) have impacted on the dependent variable. In the context of this thesis, we examine if the presence of Privacy Seal (independent variable) impacts on personal information disclosure (dependent variable). Any difference observed between the treatment and control group dependent variable outcomes are deemed to have been caused by the experimental manipulation or change in

independent variable (Bryman 2012). In other words, we are testing to see if the changes to the independent variable have a causal effect on the dependent variable.

3.6.1.1 Experiment: ‘Survey’ Questions

It is worth noting that from a participant’s perspective, both the survey and experiments appear the same, that is, they both present themselves as surveys. By employing the use of control and treatment groups the survey becomes an experiment. From the participants perspective, they are not aware that control/treatment groups exist and are unaware they are participating in an experiment. The MTurk software only allows each worker to participate in the experiment once and this means participants are not aware of independent variable manipulations that take place (or whether they are in a control or treatment groups etc).

The survey questions element used in this experiment is taken directly from a paper ‘*Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information*’ published in the Journal of Consumer Research in 2011. The authors Leslie John, Alessandro Acquisti and George Lowenstein are leading privacy research academics and this experimental research study that examined the impact contextual clues have on personal information disclosure. The questions are shown below. The survey questions are deliberately intrusive so as to,

“affect the divulgence of privacy-relevant information, which we operationalize by the sensitivity of the information that participants are asked to divulge. Information on one’s food preferences, for example, is inherently less sensitive, and hence less privacy relevant, than information one one’s sexual preferences” (John et al 2011, p.860).

By asking the participants to answer more intrusive questions about themselves, we assert that they will give greater consideration on whether they choose to disclose (or otherwise) their sensitive personal information. Please see the section below for more details.

Figure 3-7 Survey questions taken from John et al, 2011

Item
1. Have you ever made up a serious excuse, such as grave illness or death in the family, to get out of doing something?
2. Have you ever looked at pornographic material?
3. Have you ever used sex toys?*
4. Have you ever smoked marijuana (i.e., pot, weed)?
5. Have you ever "cheated" while in a relationship?
6. Have you ever driven when you were pretty sure you were over the legal blood alcohol level?
7. Have you ever taken nude pictures of yourself or a partner?*
8. Have you ever encouraged someone to drink when you were trying to seduce them?
9. Have you ever tried to peek at someone else's (e.g., a classmate's, boyfriend's, girlfriend's) e-mail account without them knowing?*
10. Have you ever fantasized about having violent nonconsensual sex?
11. Have you ever tried cocaine?
12. Have you ever had sexual thoughts about a member of your same sex?*
13. Have you ever sold marijuana (i.e., pot, weed) to someone?
14. Have you ever watched someone while they undressed, without their knowledge?*
15. Have you ever had anal sex?

(Source: John et al. 2011, p.863)

Respondents are required to answer all 15 questions presented by selecting either 'yes' or 'no' from a drop down menu (for each question). When they have answered all 15 questions the select the <next> button and they see a page that requests some demographic information from them. Please see section below for more details.

3.6.1.2 Experiment: Demographic Questions

Having completed the survey element of the experiment, the participants are asked for '*Please tell us a little bit about yourself*'. In this section they are asked for 7 pieces of personal information. For all questions, respondents are required to select and answer from the drop down menu. All questions include the option to answer the question by selecting the 'would rather not say' response.

The first three pieces of information requested are gender, age and level of education achieved. Questions 4-7 request additional pieces of personal information. The data requested here are the categories of data the greatest number of respondents identified as 'sensitive personal data' in the survey

carried out before this experiment was conducted (as described in earlier sections).

The four 'sensitive personal data' categories represent the core test for causality in this experiment. All experiment participants are presented with exactly the same survey questions and demographic questions. The only difference between the control and treatment groups is the manipulation of the independent variable. Please see section below for more detail on how the independent variable is manipulated between conditions.

This research assumes that experiment participants would give greater consideration to the decision on whether to disclose personal information (or not) if the data they were being as to disclose was considered to be sensitive personal data as apposed to 'normal' personal data. This would in turn help us to establish if the presence of the independent variable in the form of a Privacy Seal offered assurances (or otherwise) that impacted on disclosure.

3.6.1.3 Independent and Dependent Variables

A key part of any experimental design is the independent and dependent variables. Thorough the application of appropriate statistical techniques we can establish if a causal relationship exists between them and in turn answer the overarching research question.

Independent Variable:

An independent variable,




“is one that influences the dependent variable.....To establish causal relationships” (Sekaran 1992, p.66)

between it and the dependent variable. As the overarching research question this in this thesis examines, ‘*What impact(s) do changes to data protection regulations have on personal information disclosure in online environments?*’ The independent variable in the experiments conducted here is therefore an

image. Please see table below for details on the actual privacy seal images used in the treatments in the experiments. The table below describes the audiences and treatments applied:

Figure 3-8 Independent Variable display by Experiment Groups

Group	Jurisdiction	Image Displayed
Control	EU and US	No Image Displayed
Treatment_1	US only	ePrivacy Seal 
Treatment_1	EU only	Europrize Privacy Seal 
Treatment_2	EU and US	BU_logo_image

		
Treatment_3	US	<p>ePrivacy Seal and BU Logo</p> 
Treatment_3	EU	<p>Europrize Privacy Seal and BU Logo</p> 

Dependent Variable(s):

A dependent variable is,

“A variable that is causally influenced by another variable (i.e. and independent variable)” (Bryman 2012, p.711).

As the overarching research question address in this thesis examines how privacy seals impact on personal information disclosure, we are examining how the manipulation of the independent variable (i.e. privacy seal image) impacts on our dependent variables (i.e. personal data).

The expectation is that experiment participants will give greater consideration as to whether they disclose personal data that they consider sensitive rather than data they consider not to be sensitive. To answer RO2 a survey was carried out where respondents to indicate across a range of categories what data they consider as sensitive data. The survey results are used to identify the categories of personal data the highest number of respondents (in relevant jurisdiction) indicates as 'sensitive personal' and these are the dependent variables we use in the experiment.

The survey included six categories of data that are classified as 'sensitive personal data' under current EU data protection regulations. These are; 1) racial or ethnic origin, 2) political opinion(s), 3) religious and/or philosophical beliefs, 4) trade union membership, 5) health or medical history and 6) sexual orientation. These categories were initially set out in the 1950 ECHR and have been included in 95/46 EU data protection regulations and the GDPR.

The survey also includes an additional five categories of personal data that were discussed in various US and EU white papers that preceded the publication of the final draft of the GDPR. These are; 7) physical location or movement, 8) date of birth, 9) home address postal (zip) code, 10) current/past employment details and 11) personal income details.

3.6.1.4 Control and Treatment Groups

Control and treatment groups play a key role in experimental design. The control group is the group of respondents who does not receive any treatment and is used as a benchmark that we measure the results from the treatment group. In our experiment design the independent variable in the control group is blank i.e. no privacy seal image is displayed to experiment participants in the control group. Treatment groups are identical to the control group with the exception of the experimental treatment. In the experiments conducted here, the treatment involves displaying an image (i.e. privacy seal).

The independent variables manipulated in the experiments are images. These are the only changes between the control and treatment conditions. All 'business ethics' and demographic questions asked are the same, including background colours and instructions given to participants. There are some very minor changes from UK to American English and the use of the € and \$ currency symbols in accordance with relevant jurisdictions.

Control Group -> No image

In the control group the independent variable is blank, that is to say no privacy related image (or any other logo or seal) is displayed in the header and/or footer of the pages displayed to the experiment participant.

This is our baseline condition. We compare the treatments described below against this control condition and test for statistical significance.

Treatment_1 -> Privacy Seal image

In the treatment_1 condition, a privacy seal is displayed in the header and footer of each page displayed to the experiment participant. EU and US experiment participants see different privacy seals. However, both privacy seals are currently in use commercially through accrediting organisations. Written permission was granted by both seal accreditation schemes to use their images in these experiments. This is a key treatment for this experiment. Does the presence of a privacy seal impact on participant's personal information disclosure?

Treatment_2 -> Bournemouth University (BU) Logo

In the treatment_2 condition we display the same BU Logo image in EU and US jurisdictions.

This treatment condition is included to test if the presence of any logo has the same impact on the dependent variable as the Privacy Seal(s) used in treatment_1 (as described above).

Treatment_3 -> Privacy Seal and Bournemouth University Logo

In the treatment_3 condition we display the relevant Privacy Seal and BU Logo image together.

This condition is included to see if the presence of a privacy seal in conjunction with the BU logo has any impact on the dependent variable.

3.7 Data Analysis

As outlined in the previous sections, we conduct a number of surveys and the results of these then inform the dependent variables in the online experiments conducted. The treatments manipulate the independent variable and examine the impact it has on our dependent variable. The following sections outline the analysis techniques we use arrive at our research findings and the implications that follow from that. Broadly speaking our data analysis is broken down into 1) descriptive statistics and 2) inferential statistics. The following sections describe these in more detail.

3.7.1 Descriptive Statistics

The data collected in the surveys and experiments carried for this thesis presented in later chapters using both descriptive and inferential statistics.

“Descriptive statistics are used to describe the basic features of the data in a study. They provide summaries about the sample characteristics and responses to individual survey questions. Together with tables and charts, descriptive statistics form the basis for quantitative data analysis” (Sue and Ritter 2012, p.150).

The findings chapter will provide an overview of all responses from survey and experiment participants. Where appropriate these may also be shown in graphical format including frequency distributions etc.

3.7.2 Inferential Statistics

The summary descriptive statistics discussed above provide a useful overview of individual survey/experiment responses data. However, to answer the research objectives we also need to examining how two or more variables interact with each other i.e. do male/female respondents answer questions in a particular way or do respondents who have achieved a certain level of education respond a certain way? When we do this we are no longer simply

describing the data, we are making inferences from the data (Sue and Ritter 2012).

The experiment results are analysed using relevant inferential statistical methods (i.e. Fisher Exact Test detailed below) to see if the results are statistically significant when compared to our baseline control group.

“The level of statistical significance is the level of risk that you are prepared to take that you are inferring that there is a relationship between two variables” (Bryman 2012, p.348).

In other words, it tells us how confident we are that the treatment(s) have a causal effect on user information disclosure. In social sciences, the maximum level of statistical significance is generally a p-value < 0.05 which implies that there is a less than 5 chances in 100 that your sample shows a relationship where there is in fact not one.

The treatment(s) represent the ‘effect’ and care is taken to minimise the effects so as not to introduce confounding variables i.e. manipulations other than the desired effect that may influence the participants responses. For each of the experiments undertaken in this research, the effects and response variables are clearly identified.

To answer our RO3, we examine if there is a relationship between the presence of a Privacy Seal (independent variable) and whether the respondents has disclosed the relevant ‘sensitive personal data’ (dependent variable). Here we are looking at the relationship between two nominal categorical variables. By creating contingency tables and apply the appropriate statistical technique, we can establish if a statistically significant relationship exists between the two variables. In many cases a chi-squared test can be applied to test of the observed pattern in the contingency table is statistically significant or not i.e. by examining the difference between the observed and expected frequencies. There are instances where with small frequencies the chi-squared test assumptions are not met and in these cases the Fisher Exact Probability test is a more appropriate test.

The Fisher Exact test is used in the analysis of contingency tables to test for statistical significance and although it is generally used for small sample sizes, it is valid for all sample sizes (Fisher 1922).

Unlike many other statistical tests that use a mathematical function to estimate the probability of a test statistic, the Fisher Exact Test calculates the probability of getting the observed data and compares this to your treatment data to see if the differences between the groups are statistically significant (McDonald 2014).

3.8 Validation of Data

In all original research, it is important to validate whether the data and measures applied is valid.

“When we ask a set of questions (i.e., develop a measuring instrument) in hopes that we are tapping the concept, how can we be reasonably sure that we are measuring the concept we set out to measure and not something else? This can be determined by applying certain validity tests” (Sekaran 1992, p.171).

There are a number of areas where validity needs to be considered and these include internal/ external validity as well as specific validity tests. The following sections address these areas.

3.8.1 Internal Validity

Internal validity is especially relevant to research that attempts to establish causal relationships i.e. were we are trying to establish the manipulation/treatment that caused the change in outcomes. Internal validity means that you have evidence that the intervention caused the change in outcomes. Establishing internal validity is very much relevant to the experiments in this thesis.

Rosenthal and Jacobson (1968) highlight a number of threats to internal validity. They believe historical events in the wider environment may have impacted on participants' attitudes (in our case attitudes to Privacy Seals) and this may affect how participants perform in the experimental conducted. If historical events are an issue then this may that the outcome (i.e. dependent variable) is being influenced by events outside of the experiment settings. However, the presence of a control group combined with random assignment mitigates such concerns, as we would expect both the control group and any treatment group(s) to include participants that have being influenced by any such history. Any differences between treatment and control groups can be attributed to independent variable manipulation and not as a result of historical influences.

Other potential issues include the possibility that experiment participants have become sensitized to the aims of the experiment from participating in a pre-test is another area identified as a potential threat to internal validity. However, the experiments conducted for this research do not employ pre-tests. Additionally, the MTurk application does not permit workers to participate in the same survey/experiment (or HIT to use the MTurk term) more than once. Other potential internal validity issues identified include what the authors call Mortality. This is where studies span a long period of time and participants may leave the study. Maturation is also a potential issue. This recognises that people and their views change over time and this impacts on the dependent variable in ways that cannot be attributed to the independent variable. Selection, where a non-random process means that the control and treatment groups have structural differences is also a threat to internal validity.

The random assignment of participants to control and treatment groups combined with the fact that the experiments are conducted post-test only means that all of the internal validity risks discussed above are mitigated. However, it is worth noting that just because a study is considered to be internally valid it does not mean that it is beyond reproach and cannot be questioned. For example, it would be ask if independent variable manipulation valid? Do the privacy seals displayed actually look like privacy seals? Are the dependent variable(s) a valid measure of what the research sets out to address? These issues are addressed in the sections below.

3.8.2 External Validity

External validity is concerned with,

“Whether the results of a study can be generalized beyond the specific research context in which it was conducted” (Bryman 2012, p.711).

This is important as it tells us to what extent the results of this study can be generalized to other situations and to the wider population.

Cook and Campbell (1979) identify a number of threats to external validity and these are relevant to experimental research. The first area of concern is the 'Interaction of selection and treatment' i.e. are a large portion of the participants from a particular gender, race/ethnicity, social class, income group, educational achievement etc? The demographic details of participants in the survey and experiments conducted for this research will help to identify these types of issues. In our research analysis chapter and demographic issues are addressed and any limitations in relation to external validity are addressed.

'Reactive effects of experimental arrangements' is an issue where participants are aware they are participating in an experiment and this may in turn affect their responses. It is also linked with social desirability. In the experiments conducted here, the 'survey' component was deliberately described as a survey examining 'business ethics' so as not to minimise self-selection bias and potential social desirability issues. Additionally, participants are not aware that they are participating in an experiment and this combined with the fact that there is no pre-test means that this potential external validity issue is successfully mitigated. However, the possibility also exists that the experiment participant is blind to the treatment effect i.e. they do not see the logo/seal presented to them. This issue was considered in the experiment design. Highlighting the presence of the privacy seal to participants had the potential to trigger privacy concerns and induce potential self-selection bias. When running limited pilots for the experiments to test if they worked correctly, the pilot participants confirmed that they had seen the privacy seal.

'Interaction of setting and treatment' is another concern highlighted by Cook and Campbell (1975). How confident are we that the results from this experiment conducted here are applicable to other similar situations? In the context of this research one could ask, would experiment participants react in the same way in terms of disclosure of personal information if they were purchasing something online? This will be considered in greater detail in later chapters.

As with internal validity, the effect of any pre-testing in terms of participants becoming desensitized to the topic being examined is something that can

impact on external validity. However, there was now pretesting in the experiments conducted here so this is not an issue in terms of external (or internal) validity. A further threat exists in terms of whether the experiment findings can be generalized to the past and the future? If we ran the same experiment in 5 years time or 5 years ago would the results be the same.

3.8.3 Validity Tests

Validity tests can be grouped under three broad headings: content validity, criterion-related validity, and construct validity.

‘Face validity’ is the most basic index of content validity and it is essentially an intuitive process. Kidder and Judd (1986) cite an example where researchers design a test to measure degrees of speech impediment amongst a group of individuals. They took the step to have the test evaluated by a group of professional speech therapists who deemed it was valid.

Criterion-based validity comprises both concurrent and predictive validity. Concurrent validity relates to the criterion on which individuals are known to differ in relation to the concept in question. Bryman (2012) cites absenteeism as an example. We would expect those who are satisfied with their jobs to be absent from work less often than those who are less satisfied with their job. Should research findings show no difference in levels of job satisfaction amongst workers who are more frequently absent then this may cast doubt on the actual measure of job satisfaction used. Predictive validity is similar but adopts a future criterion measure rather than a current. The limited research available in this subject area makes concurrent and predictive validity difficult to measure/estimate for this research.

Construct validity,

“testifies to how well the results obtained from the use of the measure fits the theories around which the test is designed” (Sekaran 1992, p.173).

This is assessed through both convergent and discriminant validity. Convergent validity is established when the scores obtained by different instruments that measure the same concept are highly correlated. Discriminate variability is the opposite of this, when we expect the instrument to be uncorrelated.

As we have seen there are a number of internal and external threats to validity and we will revisit these again in the later discussion chapters.

3.8.4 Social Desirability Bias

“Social desirability and political correctness can often lead to respondents to give the ‘right’ answer rather than the real or valid answer to a survey question” (Sue and Ritter 2012, p.53).

This is driven by the desire to conform to social normal. Survey respondents generally give more honest answers to online surveys than they do in face-to-face interviews. However, this is not to say that surveys taken online are immune to issues related such bias.

Techniques to reduce social desirability bias include giving respondents assurances/promises regarding the anonymity and confidentiality of their data. However, as previously stated and consistent with existing research in this area, the surveys and experiments conducted here do not give explicit assurances regarding anonymity or confidentiality. This is done so as not to trigger privacy concerns and reduce self-selection bias. However, participants are told that the surveys and experiments are being undertaken for academic purposes.

Phillips (1973) outright condemns social science research on the basis of such bias. Others argue that awareness of the potential of such issues has led to measures to limit the impact of such issues (i.e. such as approval rating on MTurk etc).

3.8.5 MTurk Data Validity

Paolacci et al (2010) identify two major areas of concern in relation to the reliability and generalizability of data generated using MTurk workers. Concerns include; 1) whether MTurk workers are representative of the population as a whole; and, 2) concerns of the quality of the data that MTurk workers provide.

MTurk does not make demographic information about their workers available publically. However, a number of academic studies address the issues of uncertainty around both the demographic and data quality of workers. Paolacci et al (2010) conducted a comparative study, recruiting subjects for their experiments from three different sources, 1) MTurk, 2) a large Midwestern US university and 3) visitors to an online discussion board. All experiments were run using Qualtrics survey software. The HIT was only visible to workers who had at least a 95% approval rating and were based in the US. For the university sample, the experiment was conducted in an experimental lab setting and subjects were recruited from a pre-existing 'introductory subject pool' at the university. Discussion board subjects were recruited by posting a link to the survey on online discussion boards that host psychology experiments.

The findings examined a range of areas including respondents' demographics, non-response rates, attention, numeracy ability and performance in the actual experimental tasks. Overall the findings, "confirm that Mechanical Turk is a reliable source of experimental data", and that results,

"obtained in Mechanical Turk did not substantially differ from results obtained in a subject pool at a large Midwestern University. Moreover, response error was significantly lower in Mechanical Turk than in Internet discussion boards" (Paolacci et al 2010, p.416-417).

A paper by Rand (2012) sought to explore self-reported demographics of MTurk workers. The author posted a HIT where workers were asked to complete a short demographic questionnaire that included a question about the subjects country of residence. As MTurk automatically logs (and makes available to requesters) the IP address of each respondent, the author was able to check if the self-reported country of residence was the same as what their IP address

resolved to. Of the 176 respondents, the author found 97.2% self-report accuracy.

While the IP address was a convenient way to verify respondents' country of residence, it was not possible to verify the other self-reported demographic data with the same level of accuracy. However, the author did attempt to gain some insights into the reliability of such data by examining the consistency of workers data across different studies. Two studies were identified, each with in excess of 1,000 MTurk responses, where 100 MTurk workers had participated in both studies. The author found a high level of consistency across demographic information such as age, country of residence, education level, income level, and religious beliefs. The author concluded that the results,

“clearly indicate that most subjects are not merely making random selections” (Rand 2012, p.176).

Another strategy used by MTurk requesters to try and improve data quality is the use of attention check questions (ACQ's). A series of studies show how ACQ's can be effective at screening out inattentive MTurk workers and thus improve data quality (Aust, Diedenhofen, Ullrich and Musch 2013; Buhrmester, Kwang and Gosling 2011; Downs, Holbrook, Sheng and Cranor 2010; Oppenheimer, et al 2009). ACQ's are generally inserted in tasks as either a “trick” question(s) or an instruction requiring the respondent to answer a specific question(s) in a certain way. MTurk workers not paying sufficient attention to the HIT they are participating in (i.e. who fail to answer the ‘trick’ question(s) as directed by the requester) can be ‘filtered out’ of results (Peer, Vosgerau, and Acquisti 2014).

An alternative way of maintaining data quality is to use the information the MTurk application stores about workers based on their performance on previous tasks. Each time a worker completes a task, the requester has the option to approve or reject the workers submission (this decision also dictates whether they get paid or not). Over time workers build up their reputation (or approval rating) based on the number of approved/rejected tasks completed. Requesters can require that only workers with an approval rating above a certain

percentage (i.e. 95%) and/or only workers that have successfully completed a certain minimum threshold of tasks to participate in their HIT.

Research from Peer et al (2014) examined the ACQ and reputation based approaches used to improve MTurk data quality HIT results. Their research recruited both high-reputation and low-reputation workers to a series of HIT experiments. They also used an 'arbitrary' 95% approval rating to differentiate between high and low reputation workers. Their analysis compared the results from both groups in the context of data reliability, central-tendency bias and socially desirability of responses. The results show that high-reputation workers provided high quality data irrespective of whether ACQ's were used or not. This led the authors to conclude that,

“sampling high-reputation workers is not only a necessary, but also a sufficient, condition for obtaining high-quality data. Using ACQs does not seem to help researchers to obtain higher quality data” (Peer et al. 2014, p.1030).

Based on these findings, all surveys and experiments carried out in this research used a combination of worker reputation and number of completed tasks in order to maintain data quality.

3.8.6 MTurk: Limitations

As we have seen, there are measures requesters can take to improve the data quality collected from HITs on MTurk. However, that is not to say that this data collection method is flawless.

MTurk does not permit workers to have more than one account and the software has a default setting that only lets workers complete a requesters HIT once. This helps to reduce potential duplication of responses i.e. the same worker completing the task over and over again. However, requesters need to be aware that if they post the same task twice (i.e. if they decide they need a bigger sample), there is the potential a 'sizable fraction' of repeat participants

for the same HIT (Rand 2012). As such, requesters need to be conscious of this possibility.

3.8.7 Ethics

Bournemouth University requires all research to have ethical approval from the BU Research Ethics Committee prior to data collection. Ethical approval was sought and granted for this research project.

4 Chapter 4: Findings and Analysis

This chapter presents the key findings in relation to the survey(s) and experiments conducted for this thesis. Each of the three research objectives is addressed separately. For each research objective, a summary of the key conclusions from existing relevant literature is presented. Key findings from the surveys and/or experiments conducted in this research are then presented. An interpretation of key findings, including consistencies/inconsistencies of these findings in context of existing literature is then addressed.

We begin with a quick recap of the research objectives and their context.

4.1 Research Objective 1: Analysis and Discussion

To examine if respondents can identify the types of data categorised as ‘sensitive data’ status under current EU data protection regulations.

The 1995 Data Protection Directive (95/46/EU) is where we first see a formal establishment of ‘sensitive data’ protection in European law. The overall aim of the directive was to promote the functionality of the single market by addressing divergent data protection provisions among member states (Cate 1999). Harmonized data protection law creates a level playing field for all EU member states and this is a key goal of the single market. It is important that common laws are applied in all member states thus preventing firms from choosing one location to seek competitive advantage i.e. to take advantage of less stringent privacy rules in a particular member state.

The 1995 directive sets out a number of broad principles covering purpose limitation, data quality, data security etc. One of the principles included special protection for ‘sensitive data’. This imposed restrictions on the collection and processing of data that can be used to identify a person’s racial or ethnic origin, political opinions, religious or philosophical belief or concerning health or sexual orientation (Loring 2002). The EU Directive adopted these categories from 1950 European Convention on Human Rights which itself was created as a result of the horrific experiences in concentration/death camps in Europe during WWII.

There is no existing literature this author is aware of that specifically examines individual’s knowledge or ability to identify of categories of sensitive data under EU data protection regulations. Therefore the findings described in the sections below represent one of the original contributions to knowledge of this thesis.

4.1.1 RO1: 'Sensitive Data' Survey Findings

This section presents the findings from the survey conducted in relation to RO1. The demographics of the respondents are presented and then the response to each of the 8 x survey questions are presented.

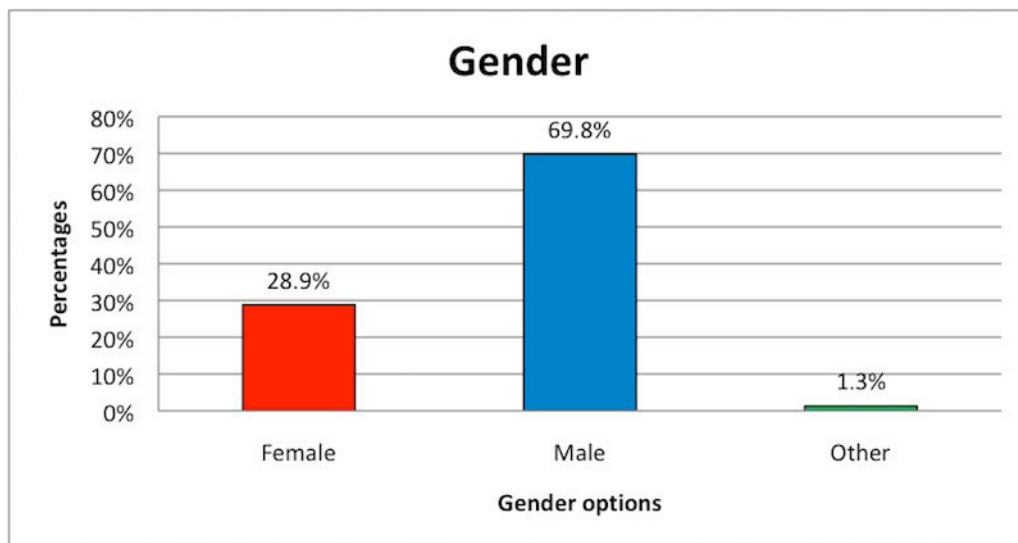
4.1.2 RO1 Survey: Demographics

This section presents the basic demographics of those who participated in the survey including gender, age and level of education in relation to survey the MTurk 'workers' who responded to the survey.

4.1.2.1 RO1 Survey: Gender

Respondents were asked: "What gender are you?"

Figure 4-1 RO1 - Gender (EU)



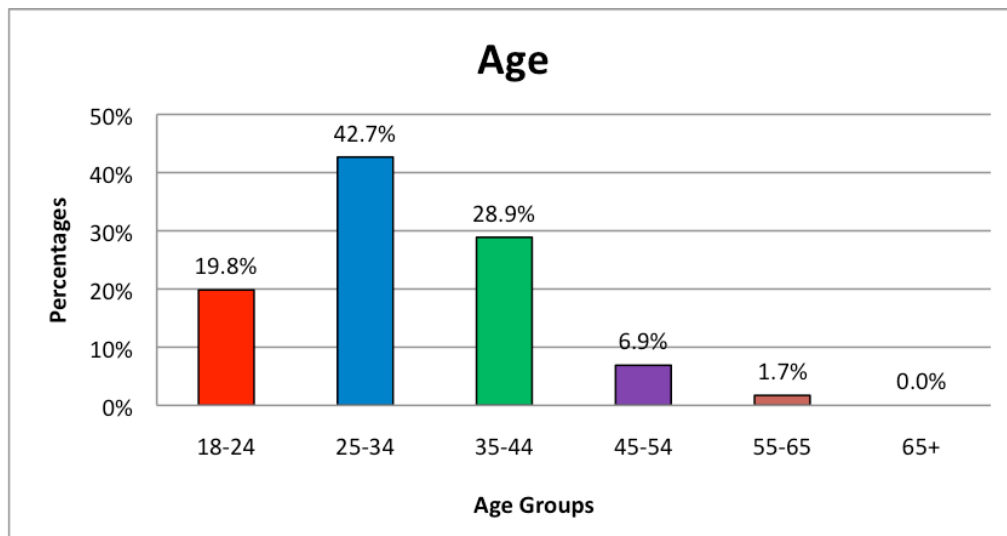
As we can see from the chart above, male participants represented 69.8% of 232 valid responses while females represented 28.9%. Respondents who selected 'other' were just 1.3% of respondents.

4.1.2.2 RO1 Survey: Age

Respondents were asked: “*What age are you?*”

A drop down menu required the respondent to select from one of six options, they were: a) 18-24 years old b) 25-34 years old c) 35-44 years old d) 45-54 years old e) 55-65 years old and f) 65+ years

Figure 4-2 RO1 – Age (EU)



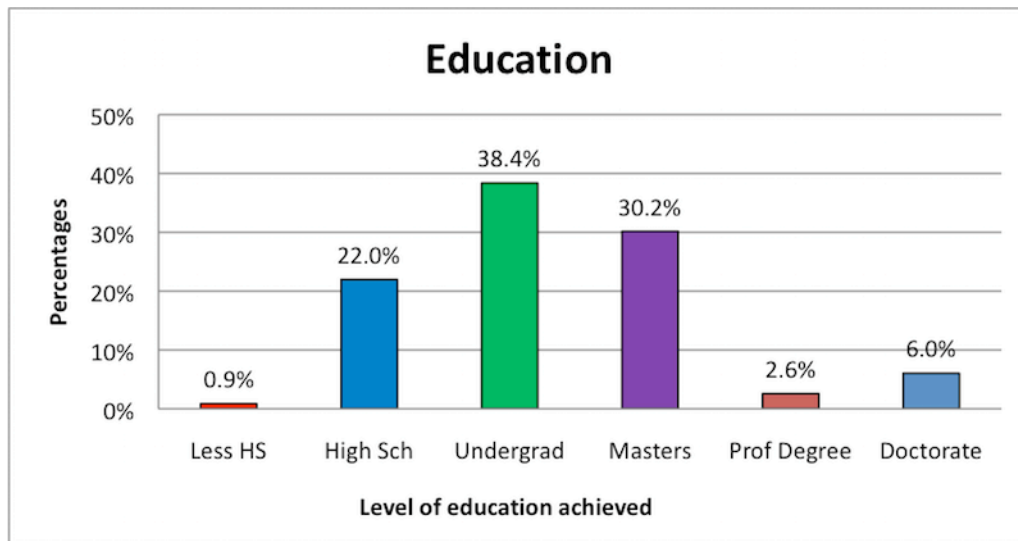
Out of the 232 valid survey responses, the 25-34 year age range were the largest group representing 42.7% of all respondents. Next was the 35-44 age range making up 28.9%. The third largest group were 18-24 year olds at 19.8%. The 45-54 age range constituted 6.9% of the sample while 55-65 year range made up just 1.7% of respondents. There were no respondents in the 65+ age range.

4.1.2.3 RO1 Survey: Education

Respondents were asked: “Please indicate your level of education?”

A drop down menu required the respondent to select from one of three options, they were: a) Less than secondary/high school b) Secondary/high school completed c) Undergraduate degree completed d) Masters degree completed e) Professional Degree completed f) Doctorate completed

Figure 4-3 RO1 – Education (EU)

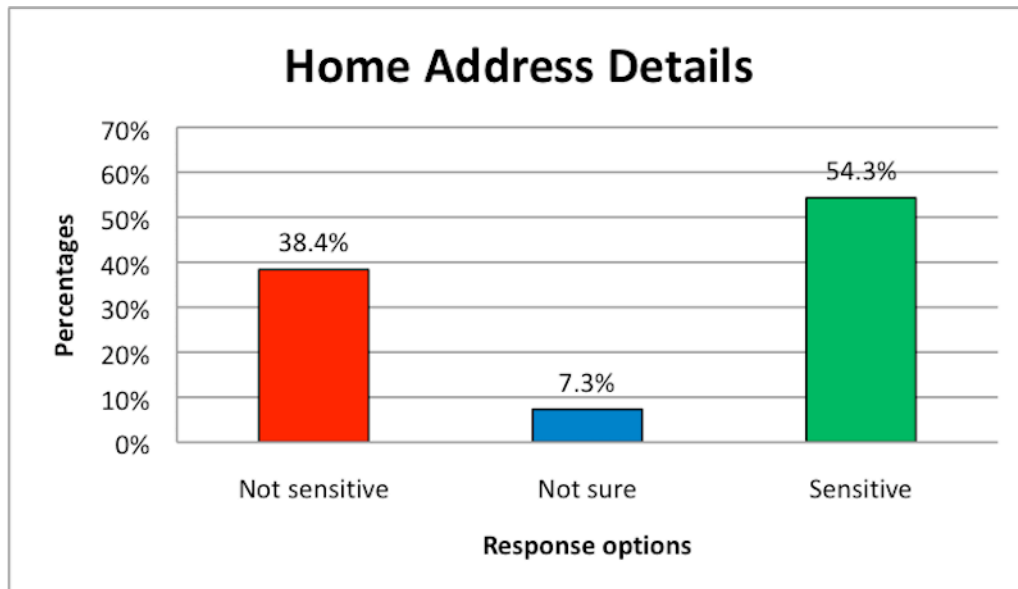


As we can see, 99% of respondents had at least finished high school with 77.1% having completed an undergraduate degree.

4.1.2.4 RO1 Survey: Home Address Details

Respondents were asked: “Do you believe details that identify your **home address details** (i.e. postcode/zipcode) are categorized as 'sensitive data' under current EU data protection regulations?”

Figure 4-4 RO1 – Home Address Details (EU)

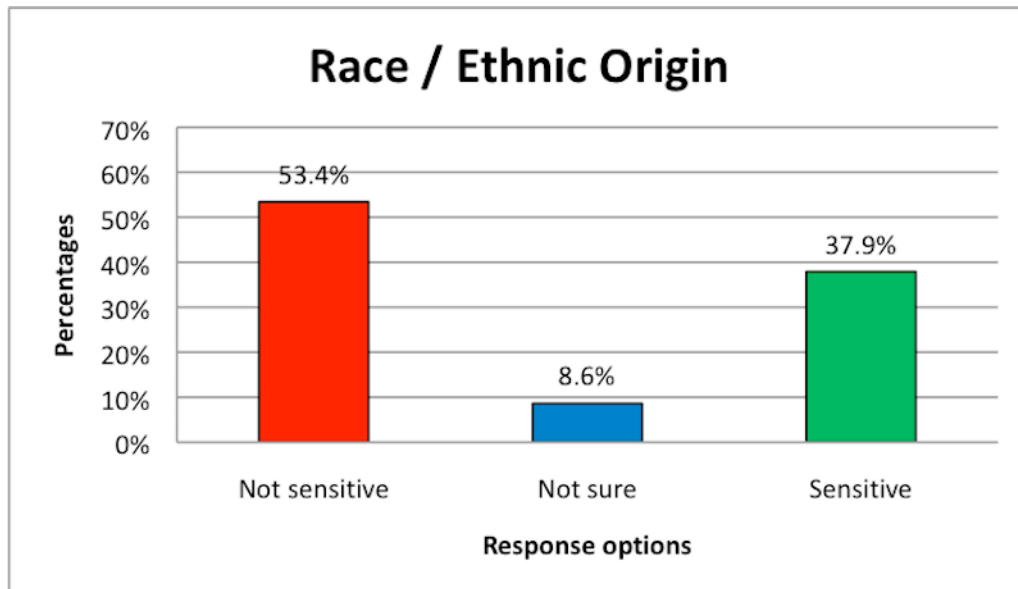


The results show that out of the 232 valid responses, 54.3% thought that details that identify your home address are categorised as sensitive data under current EU data protection regulations. 38.4% of respondents thought it was not sensitive data and 7.3% indicated they were not sure/did not know.

4.1.2.5 RO1 Survey: Race / Ethnic Origin

Respondents were asked: “Do you believe details that identify your **racial or ethnic origin** are categorized as 'sensitive data' under current EU data protection regulations?”

Figure 4-5 RO1 – Race / Ethnic Origin (EU)

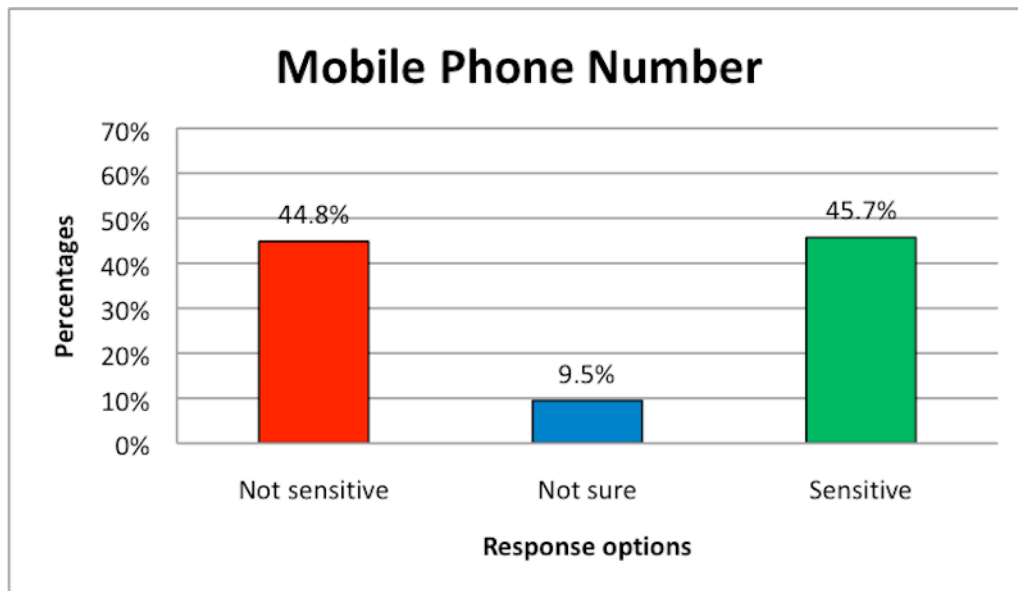


The findings show that out of the 232 responses, 53.4% believe details that identify your racial or ethnic origin are not categorised as sensitive data under current EU data protection regulations. 37.9% of respondents thought it was classified as sensitive data and 8.6% indicated they were not sure/did not know.

4.1.2.6 RO1 Survey: Mobile Phone Number

Respondents were asked: “Do you believe your mobile telephone number is categorized as 'sensitive data' under current EU data protection regulations?”

Figure 4-6 RO1 – Mobile Phone Number (EU)

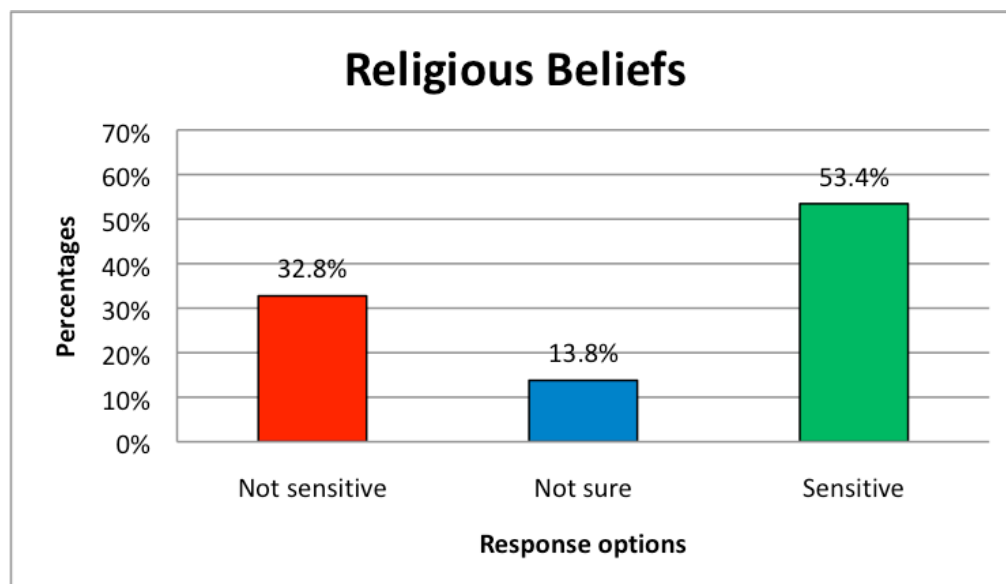


The results here were a pretty even split with 45.7% of respondents indicating that they thought that their mobile phone number was categorised as sensitive data under current EU data protection regulations, while 44.8% thought it was not sensitive data. The remaining 9.5% of respondents indicated they were not sure/did not know.

4.1.2.7 RO1 Survey: Religious Beliefs

Respondents were asked: “Do you believe that details that identify your **religious beliefs** are categorized as 'sensitive data' under current EU data protection regulations?”

Figure 4-7 RO1 – Religious Beliefs (EU)

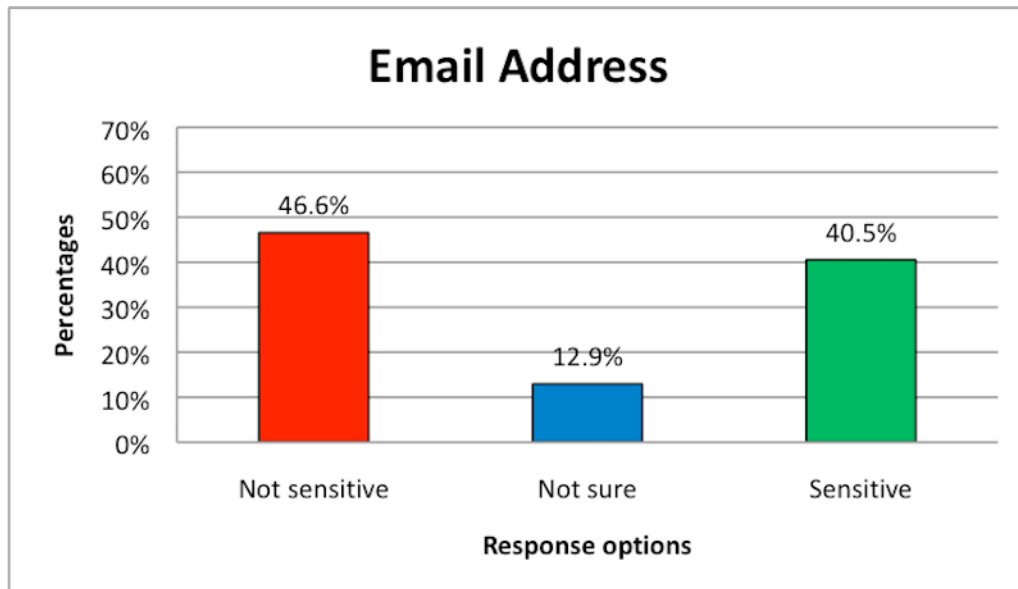


The results show that 53.4% of respondents believe that details disclosing their religious beliefs are categorised as sensitive data under current EU data protection regulations, while 32.8% thought it was not sensitive data. The remaining 13.8% of respondents indicated they were not sure/did not know.

4.1.2.8 RO1 Survey: Email Address

Respondents were asked: “Do you believe that your **email address (i.e. work or personal)** details are categorised as 'sensitive data' under current EU data protection regulations?”

Figure 4-8 RO1 – Email Address (EU)

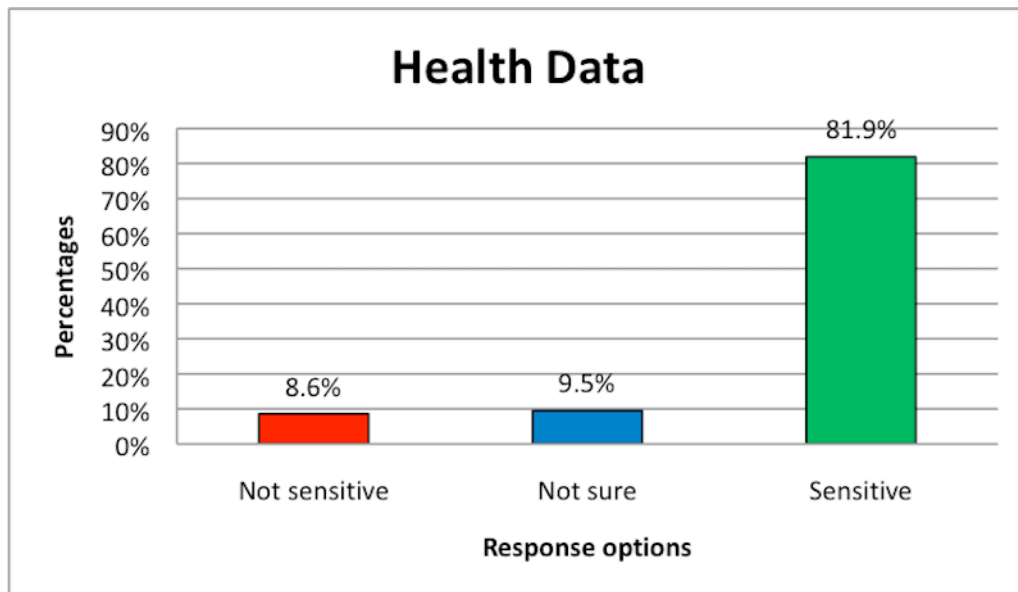


The results show that 46.6% of respondents thought a work or personal email address was not categorised as sensitive data under current EU data protection regulations, while 40.5% thought it was sensitive data. The remaining 12.9% of respondents indicated they were not sure/did not know.

4.1.2.9 RO1 Survey: Health Data

Respondents were asked: “Do you believe details that identify **any existing physical and/or mental health condition** you might currently have or have suffered from in the past are categorized as 'sensitive' data under current EU data protection regulations?”

Figure 4-9 RO1 – Health Data (EU)

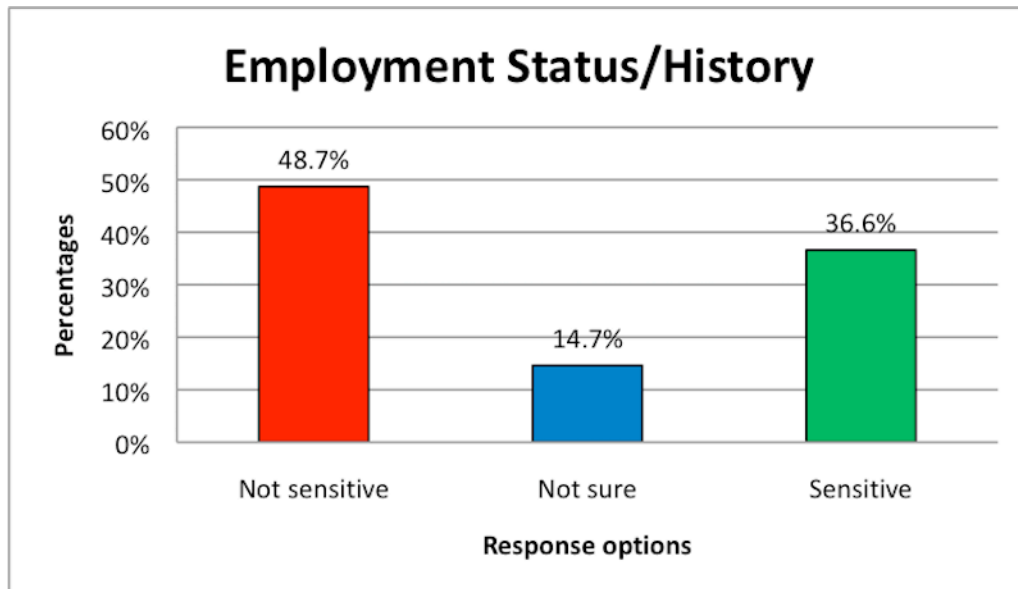


The results show that 81.9% of respondents said that they thought details relating to any existing physical and/or mental health condition is categorised as sensitive data under current EU data protection regulations, while only 8.6% thought it was not sensitive data. 9.5% of respondents indicated they were not sure/ did not know.

4.1.2.10 RO1 Survey: Employment

Respondents were asked: “Do you believe details relating to your **employment status/history** are categorised as 'sensitive data' under current EU data protection regulations?”

Figure 4-10 RO1 – Employment Status/History (EU)

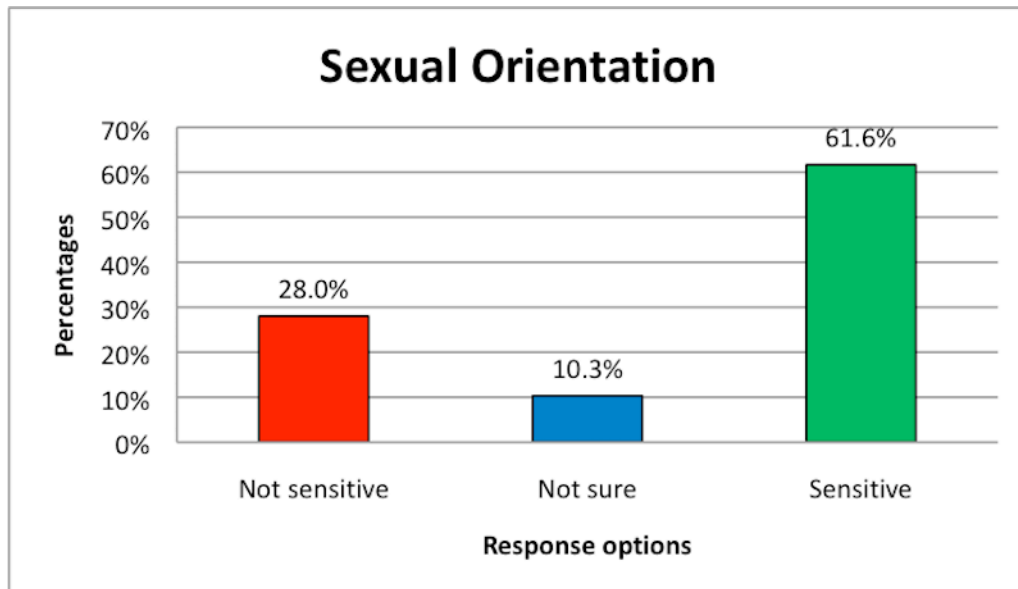


The results show that 48.7% of respondents thought details relating to employment status/history is not categorised as sensitive data under current EU data protection regulations, while only 36.6% thought it was sensitive data. The remaining 14.7% of respondents indicated they were not sure/ did not know.

4.1.2.11 RO1 Survey: Sexual Orientation

Respondents were asked: “Do you believe details that identify your **sexual orientation** are categorized as 'sensitive data' under current EU data protection regulations?”

Figure 4-11 RO1 – Sexual Orientation (EU)



The results show that 61.6% of respondents said that they thought details that identify sexual orientation were categorised as sensitive data under current EU data protection regulations, while 28% thought it was not sensitive data. 10.3% of respondents indicated they were not sure/ did not know.

4.1.2.12 RO1 Survey: Analysis & Discussion

One of the difficulties associated with a non-probability convenience sample like MTurk is that it is likely your sample data does not represent the wider general population. This is certainly true with the findings here. With 69.8% of respondents saying they are male as compared to 28.9% female respondents, the data is skewed towards male respondents. Similarly, 61.2% of respondents indicated that they had completed an undergraduate degree and this is a higher level of educational achievement than one would expect across EU member states.

It was not possible to establish if this survey sample is representative of overall registered MTurk 'workers'. Having contacted MTurk support team by email, they confirmed that they do not make information detailing age, gender or education levels of their registered workers available publically. With this in mind we now examine the results in more detail.

Below is a summary table that has been sorted (highest to lowest) by the highest number of respondents who indicated they thought that category of data was classified as 'sensitive personal data' under current EU data protection rules. The actual categories of data that are sensitive data are displayed in bold.

Table 4-1 RO1 (EU) Responses Ordered By Sensitive Percentages

Rank	Data Type	Sensitive	Not sensitive	Not Sure
1st	Physical and/or mental health	81.9%	8.6%	9.5%
2nd	Sexual orientation	61.6%	28%	10.3%
3rd	Home address details	54.3%	38.4%	7.3%
4th	Religious beliefs	53.3%	32.8%	13.8%
5th	Mobile telephone number	45.7%	44.8%	9.5%
6th	Email address	40.5%	46.6%	12.9%
7th	Racial or ethnic origin	37.9%	54.4%	8.6%
8th	Employment status	36.6%	48.7%	14.7%

Note: All figures have been rounded to 2 decimal places.

Sensitive data categories are in **bold**

The results displayed in Table 4-1 above show that three out of the top four categories that have the highest percentages of 'sensitive' responses are indeed categories of data that are afforded sensitive personal data status/protection under current EU data protection regulations. The three categories identified correctly identified are Physical and/or mental health (81.9%), Sexual orientation (63.8%) and Religious beliefs (53.3%). These findings indicate that respondents have a reasonably good understanding of what is currently classified as sensitive personal data.

However, only 37.9% of respondents indicated 'Racial or ethnic origin' as a category of personal data classified as 'sensitive', despite it having this protection under current regulations.

Interestingly, significant numbers of respondents thought that personal information that indicated their 'home address details' (54.3%), 'mobile telephone number' (45.7%) and 'email address' (40.5%) were given additional 'sensitive personal data' protections even though they are not afforded this protection in law.

This would suggest that in terms of data protection, regulations including the new GDPR, have failed to keep up with technological developments and data subjects expectations in terms of what they consider to be sensitive data.

In advance of the GDPR becoming effective in EU member states on 25th May 2018, social media service Facebook explicitly sought consent from users in order to share sensitive personal information on sexual preferences, religious views and political views. This explicit consent had not been sought under the 95 directives.

The main purpose of this survey was to establish what, if any, understanding data subjects had of the ideas of sensitive categories of personal data. The findings clearly show that out of the 8 categories of data presented, over 50% of respondents correctly identified 3 out of the 4 sensitive categories of personal data (i.e. physical and/or mental health (81.9%), sexual orientation (63.8%) and

religious beliefs (53.3%)). Only 37% of respondents thought that personal data revealing a data subjects 'Racial or Ethnic origin' was afforded 'sensitive' data protection under the EU 95/46 directive even though it is afforded such protection under the directive.

Interestingly, the results also show that significant percentages of respondents wrongly believe that some categories of data are afforded additional protection under data protection regulations when they are not. These data categories can be used/shared to target advertising messages by data controllers who users have disclosed their personal data to.

These results do provide some positives and some negatives for regulators. On a positive note, the findings show that many respondents are able to identify many of the categories of sensitive personal data as per the 95 directive. However, the findings also highlight some issues of potential concern for regulators. Significant percentages of respondents thought that categories of data were regarded as 'sensitive' personal data when they are not afforded this protection under EU data protection provisions.

4.2 Research Objective 2: Analysis and Discussion

We now address our research objective 2:

To examine what categories of personal data respondents consider ‘sensitive personal data’.

As far as this author is aware, there is no existing research that specifically examines what categories of personal data people consider/believe are ‘sensitive’ personal data or otherwise.

To address this research objective a survey is conducted. The survey presents respondents with a series of data categories and asks them whether they think the category should be classified as sensitive data (or not). As the questions in this survey do not relate to any specific piece of legislation, survey participant recruitment does not need to be limited to EU member state citizens only. The results of the survey will give us a number of important insights, including the following:

- **Are categories of ‘sensitive data’ still relevant:** As the original 6 categories of ‘sensitive’ data date from 1950, the survey findings here will examine if respondents believe these types of data should still be considered ‘sensitive’ data.
- **Identify ‘new’ sensitive data categories:** The survey asks respondents whether specific categories of personal data should be classified as sensitive personal data. The categories with the highest percentage of respondents consider ‘sensitive personal data’ will be used in the experiments conducted in RO3 (described in later sections of this chapter). The findings will provide insights into whether regulators should consider expanding the categories of data that should be considered ‘sensitive personal data’ and given additional regulatory protection.

- **Compare EU and US survey responses:** The survey participants are recruited through MTurk and are from two discrete audiences i.e. the EU and the US. As discussed in the literature review, the origins of data protection in the EU and US come from opposing philosophical approaches. The findings here will allow for a meaningful comparison of the results from the different jurisdictions. Do EU and US respondents differ in terms of what categories of personal data they consider should be classified as 'sensitive' personal data?

We address these and related issues in the following sections.

4.2.1 RO2 Sensitive Data: Survey

The 95/46 data protection directive specifically set out 6 x categories of personal data that are considered ‘sensitive’ personal data and this provision limits how data controllers can share and use these. The categories are listed in the table below.

Figure 4-12 Existing Sensitive Data Categories

Num	Sensitive data category
1.	Racial or ethnic origin
2.	Political opinions
3.	Religious and/or philosophical beliefs
4.	Trade union membership
5.	Health or medical history
6.	Sexual orientation

The survey carried out in relation to RO2 will present these 6 categories of data and examine whether respondents believe these should be considered sensitive data.

In addition to this, the survey will explore the idea of including additional categories of sensitive data. In the years preceding the publication of the final draft of the GDPR, a number of reports/white papers from EU and US Data Protection regulators proposed expanding the categories of data classified as ‘sensitive personal data’. For example, a 2010 report from the FTC discussed how any future privacy framework,

“should recognize additional harms that might arise from unanticipated uses of data. These harms may include the unexpected revelation of previously private information, including both sensitive information (e.g., health information, **precise geolocation information**) and less sensitive information (e.g., purchase history, **employment history**) to unauthorized third parties” (FT, 2010, p.8).

The report goes on to describe,

“a general consensus that information about children, **financial** and health **information**, Social Security numbers, and precise, individualized geolocation data is sensitive and merits heightened consent methods. In addition, some commenters suggested that information related to race, religious beliefs, ethnicity, or sexual orientation, as well as biometric and genetic data, constitute sensitive data” (FTC 2010, p.58-59)

Date of birth was included as recent data shows that only 3% of Facebook users disclose this information publically i.e. beyond their friends lists (Farahbakhsh, Mohammadi, Han, Cuevas and Crespi 2017) and it will be interesting to see whether respondents consider this data sensitive (or not).

Home address or postcode was included to see how respondents view this data. Many mobile phone applications use phone GPS data and this can be used to establish you home address.

We know (see above) that the FTC considered including ‘financial information’ as sensitive data so we ask respondents here how they view data related to their personal income. The survey questions will include the 5 x categories of data outlined in the table below.

Figure 4-13 ‘New’ Sensitive Data Categories

Num	Sensitive data category
7.	Physical location or movement
8.	Date of birth
9.	Home address or postcode
10.	Current/past employment details
11.	Personal income details

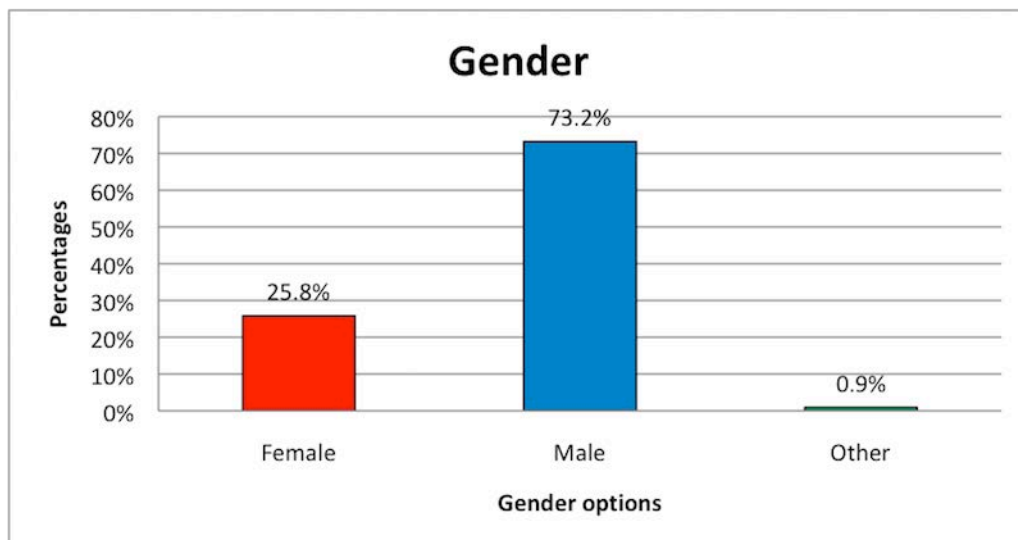
4.2.2 RO2 Sensitive Data: EU Findings

The results of the EU survey are presented in a series of tables in this section. Some basic demographics (gender, age, education etc) are presented. We then present the answers to each of the questions asked. A summary table then shows the answers to the questions indexed by highest percentage of 'yes' answers.

4.2.2.1 RO2 EU Survey: Gender

Respondents were asked: “How do you believe data revealing your **racial or ethnic origin** should be classified under data protection laws?”

Figure 4-14 RO2 - Gender (EU)

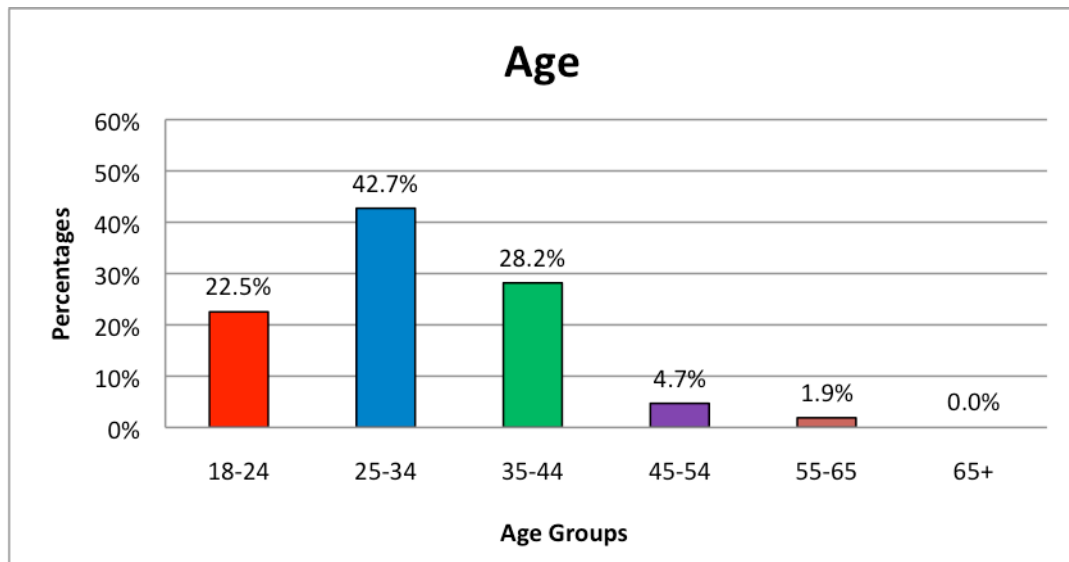


The results show that a majority of the survey participants are male and make up 73.2% of the 213 valid responses. Female respondents represented just over a quarter of the total at 25.8%.

4.2.2.2 RO2 EU Survey: Age

Respondents were asked: “How do you believe data revealing your **racial or ethnic origin** should be classified under data protection laws?”

Figure 4-15 RO2 - Age (EU)

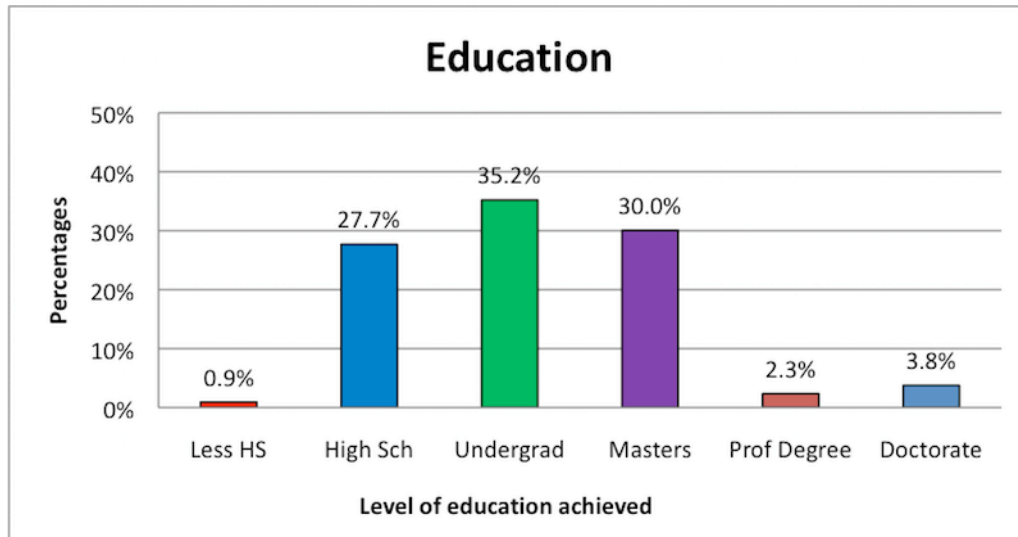


Out of the 213 valid survey responses, the 25-34 year age range were the largest group representing 42.7% of all respondents. Next was the 35-44 age range making up 28.2%. The third largest group were 18-24 year olds at 22.5%. The 45-54 age range constituted 4.7% of the sample while 55-65 year range made up just 1.9% of respondents. There were no respondents in the 65+ age range.

4.2.2.3 RO2 EU Survey: Education

Respondents were asked: “How do you believe data revealing your **racial or ethnic origin** should be classified under data protection laws?”

Figure 4-16 RO2 - Education (EU)

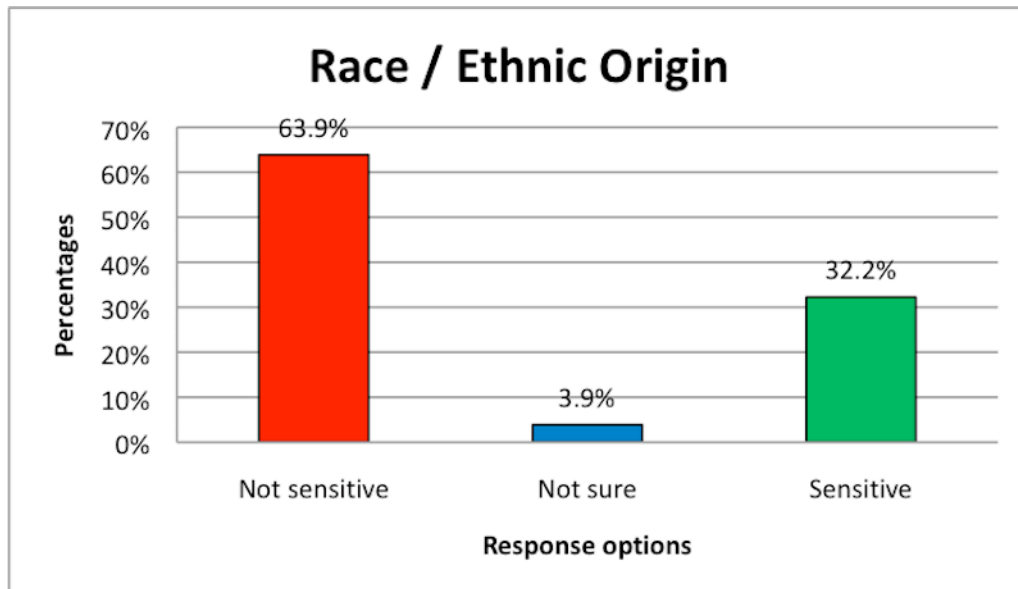


As we can see, 99% of respondents had at least finished high school with 71.4% having completed an undergraduate degree. In total 36.1% of participants said they had completed a Masters, Doctorate or professional degree.

4.2.2.4 RO2 EU Survey: Racial / Ethnic Origin

Respondents were asked: “How do you believe data revealing your **racial or ethnic origin** should be classified under data protection laws?”

Figure 4-17 RO2 – Race / Ethnic Origin (EU)

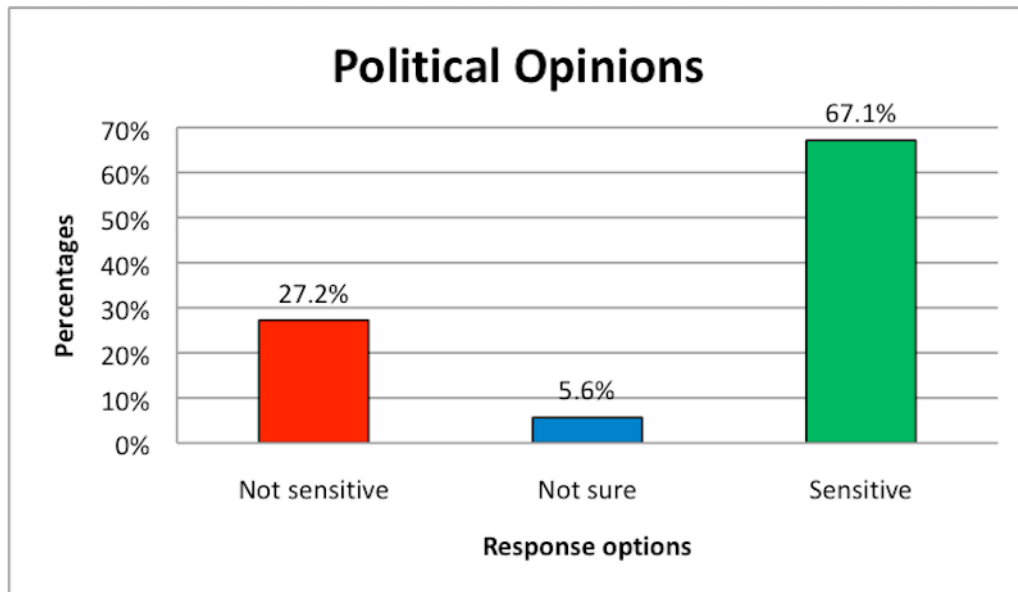


Out of the 213 valid responses, 63.9% believe details that identify your racial or ethnic origin should not be categorised as sensitive data under current EU data protection regulations. 32.2% of respondents thought it should be classified as sensitive data and 3.9% indicated they were not sure/did not know.

4.2.2.5 RO2 EU Survey: Political Opinions

Respondents were asked: “How do you believe data revealing your **political opinion(s)** should be classified under data protection laws?”

Figure 4-18 RO2 – Political Opinions (EU)

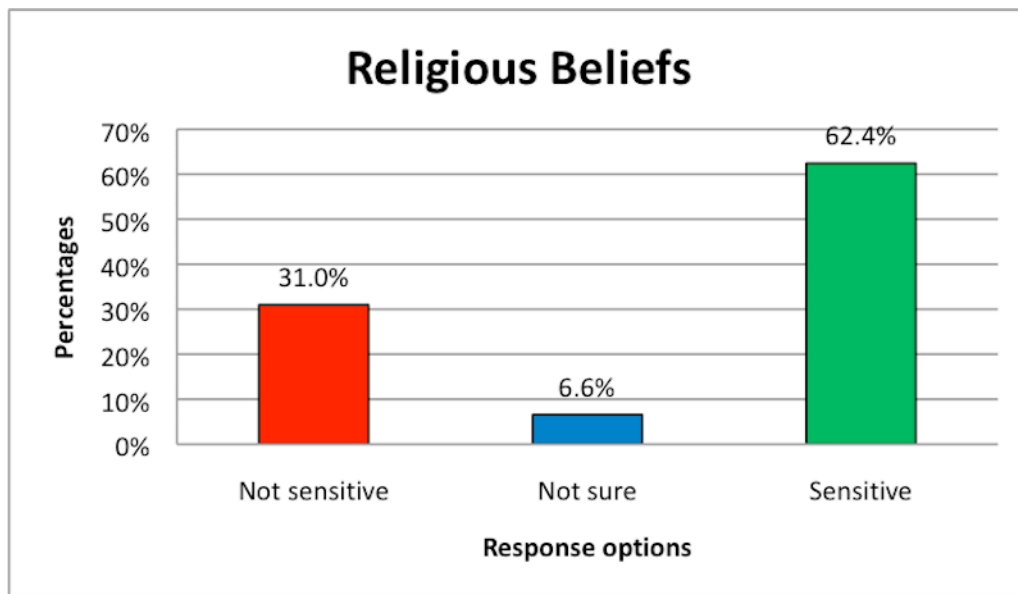


When asked about their political opinions, 67.1% of respondents thought that this should be categorised as sensitive data under data protection regulations. Only 27.2% of respondents thought it this category of personal data should not be classified as sensitive data. The remaining 5.6% indicated they were not sure/did not know.

4.2.2.6 RO2 EU Survey: Religious Beliefs

Respondents were asked: “How do you believe data revealing your **religious and/or philosophical belief(s)** should be classified under data protection laws?”

Figure 4-19 RO2 – Religious Beliefs (EU)

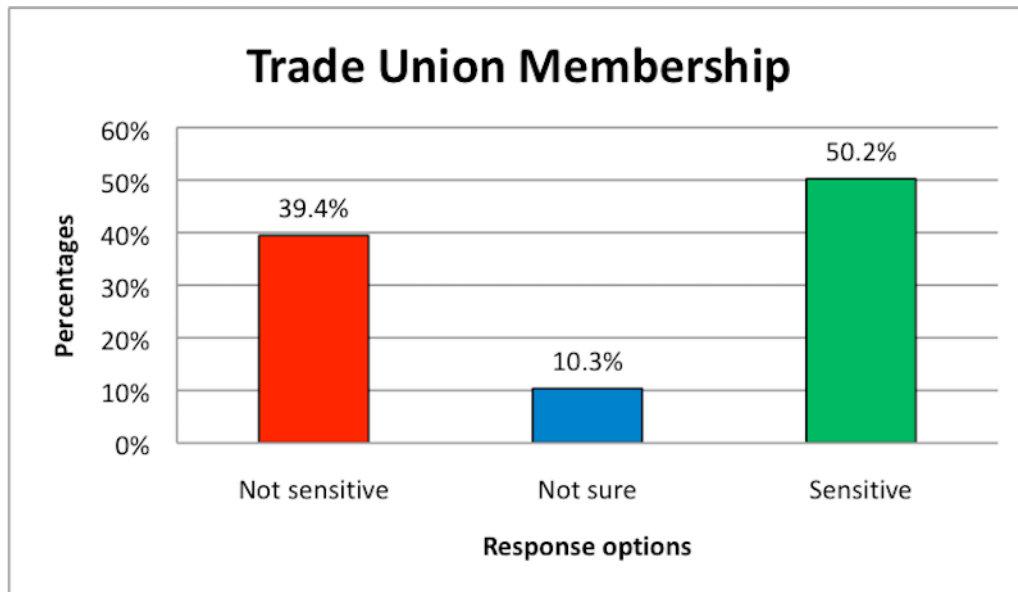


When asked about their religious beliefs, 62.4% of respondents thought that this should be categorised as sensitive data under data protection regulations. 31% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 6.6% indicated they were not sure/did not know.

4.2.2.7 RO2 EU Survey: Trade Union

Respondents were asked: “How do you believe data revealing your **trade union membership** should be classified under data protection laws?”

Figure 4-20 RO2 – Trade Union Membership (EU)

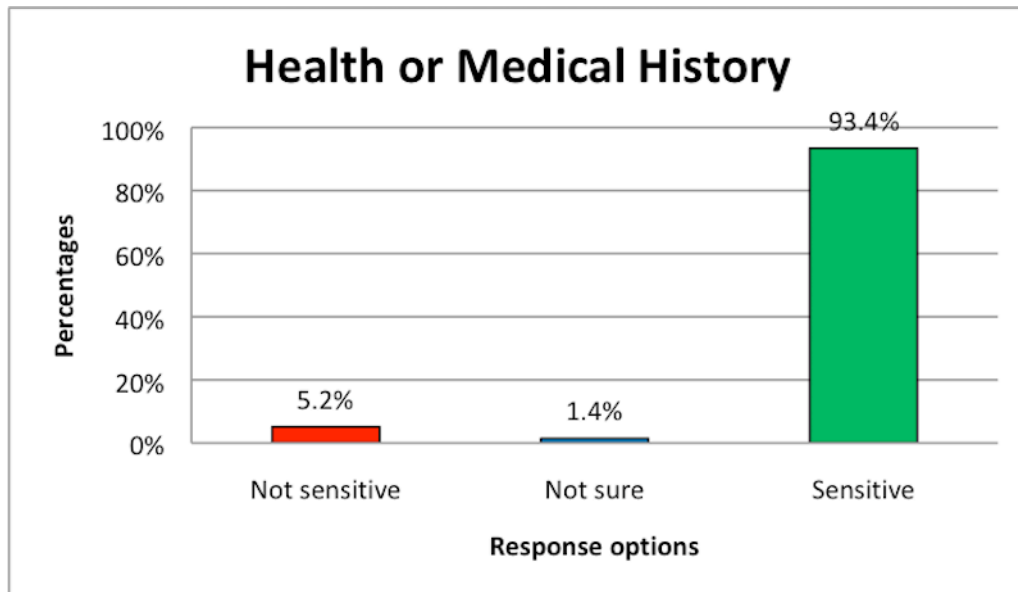


When asked about their trade union membership, 50.2% of respondents thought that this should be categorised as sensitive data under data protection regulations. 39.4% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 10.3% indicated they were not sure/did not know.

4.2.2.8 RO2 EU Survey: Health Data

Respondents were asked: “How do you believe data revealing your **health or medical history** should be classified under data protection laws?”

Figure 4-21 RO2 – Health or Medical History (EU)

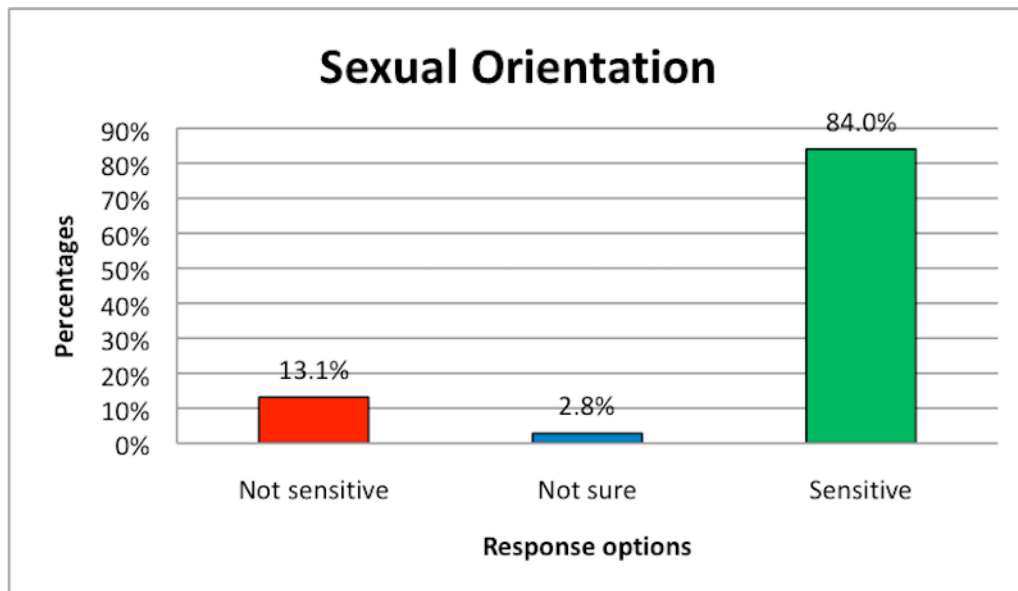


When asked about their data disclosing their health or medical history, 93.4% of respondents thought that this should be categorised as sensitive data under data protection regulations. Only 5.2% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 1.4% indicated they were not sure/did not know.

4.2.2.9 RO2 EU Survey: Sexual Orientation

Respondents were asked: “How do you believe data revealing your **sexual orientation** should be classified under data protection laws?”

Figure 4-22 RO2 – Sexual Orientation (EU)

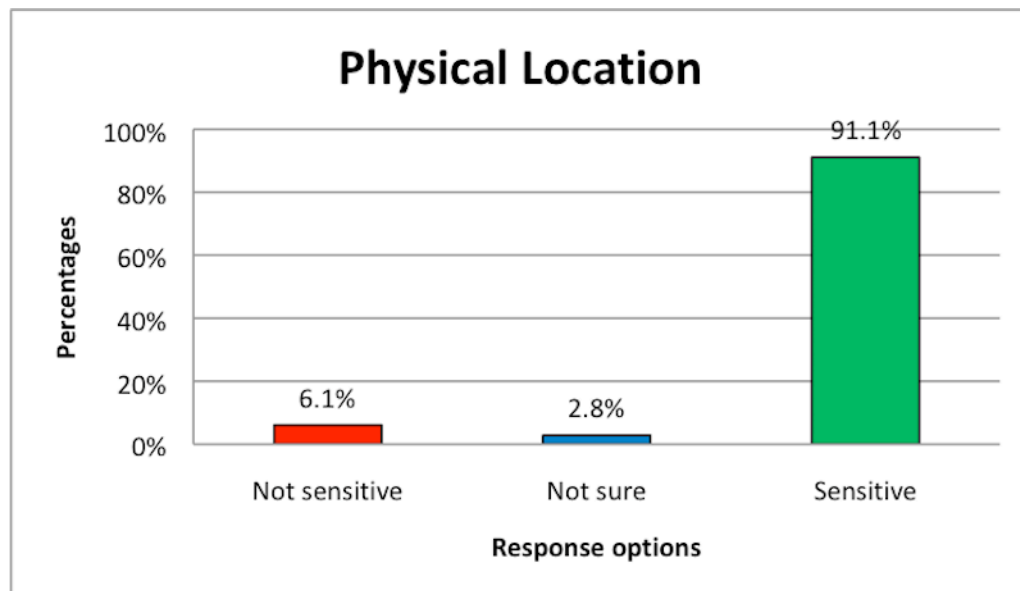


When asked about their data disclosing their sexual orientation, 84% of respondents thought that this should be categorised as sensitive data under data protection regulations. Only 13.1% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 2.8% indicated they were not sure/did not know.

4.2.2.10 RO2 EU Survey: Location

Respondents were asked: “How do you believe data revealing your **physical location or movement** should be classified under data protection laws?”

Figure 4-23 RO2 – Physical Location (EU)

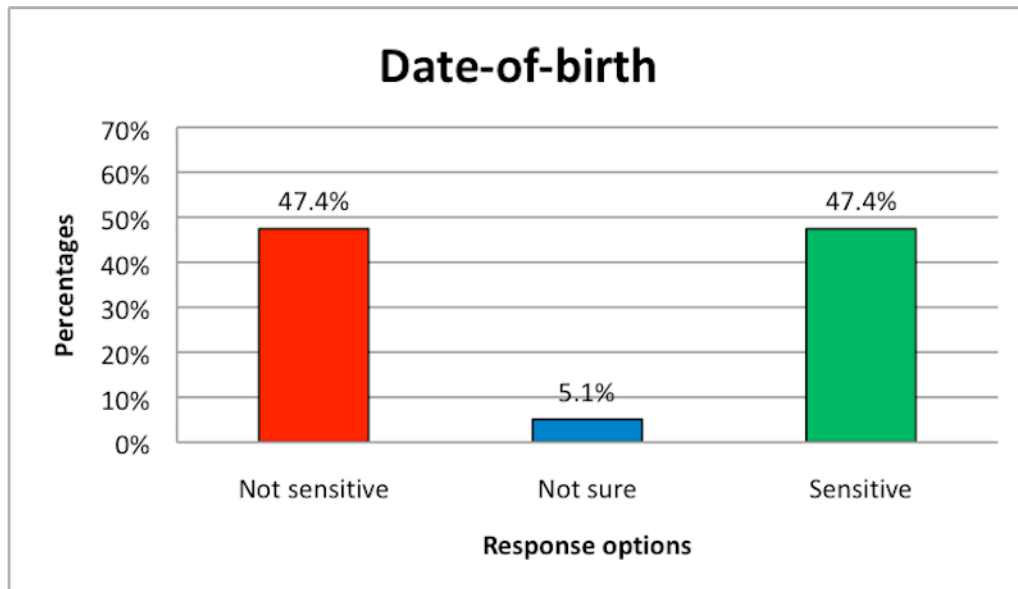


When asked about their data disclosing their physical location or movement, 91.1% of respondents thought that this should be categorised as sensitive data under data protection regulations. Only 6.1% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 2.8% indicated they were not sure/did not know.

4.2.2.11 RO2 EU Survey: Date-of-birth

Respondents were asked: “How do you believe data revealing your **date of birth** should be classified under data protection laws?”

Figure 4-24 RO2 – Date-of-birth (EU)

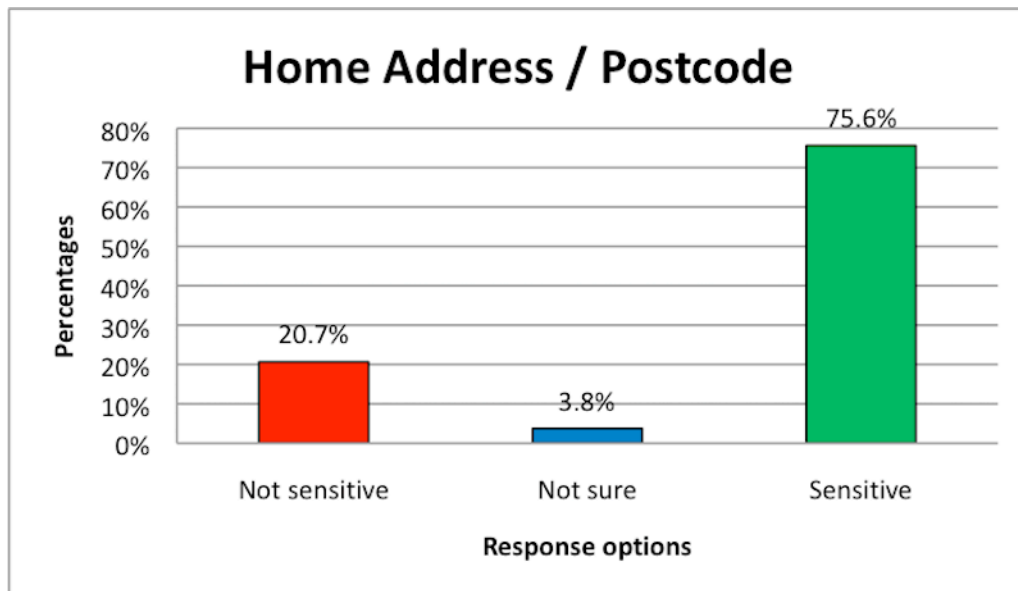


When asked about their data disclosing their date-of-birth the results were an even split with 47.4% of respondents thought that this should be categorised as sensitive data and 47.4% thought it this category of personal data should not be classified as sensitive data. The remaining 5.1% indicated they were not sure/did not know.

4.2.2.12 RO2 EU Survey: Home Address

Respondents were asked: “How do you believe data revealing your home address postal code should be classified under data protection laws?”

Figure 4-25 RO2 – Home Address / Postcode (EU)

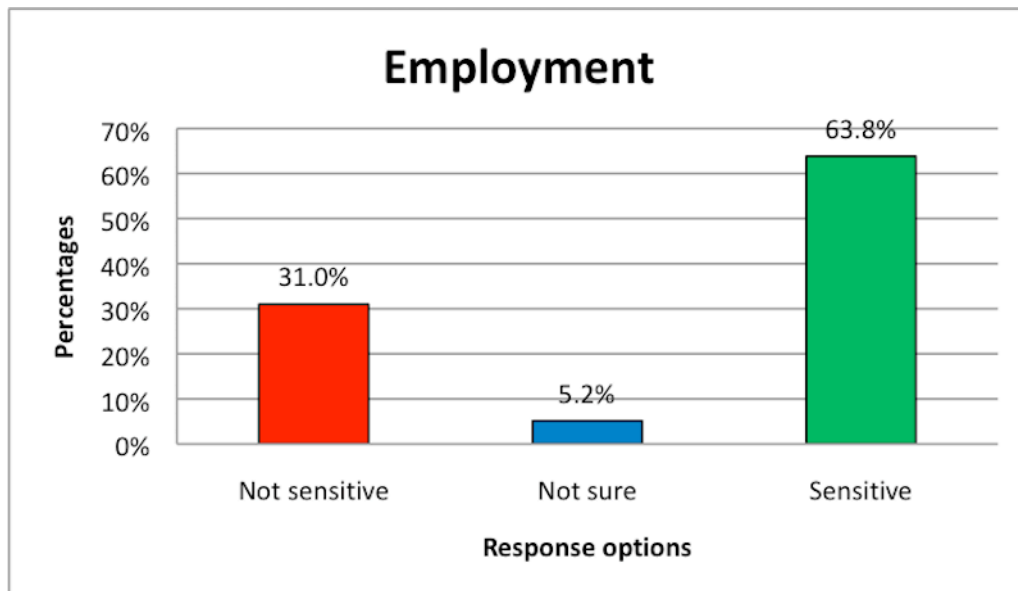


When asked about their data disclosing their how address post/zip code, 75.6% of respondents thought that this should be categorised as sensitive data under data protection regulations. 20.7% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 3.8% indicated they were not sure/did not know.

4.2.2.13 RO2 EU Survey: Employment

Respondents were asked: “How do you believe data revealing your current/past employment details should be classified under data protection laws?”

Figure 4-26 RO2 – Employment (EU)

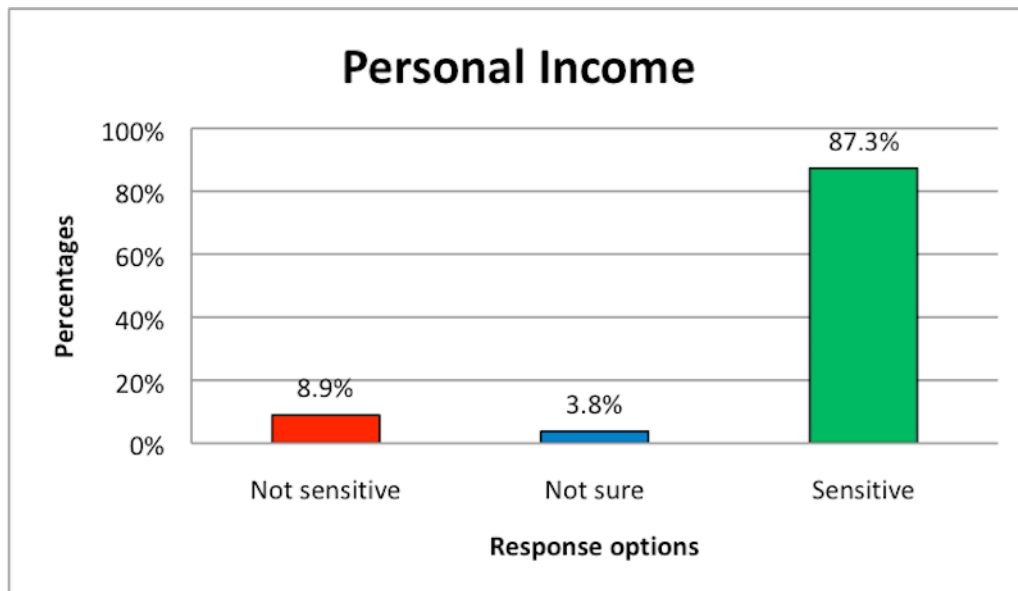


When asked about their data disclosing their how address post/zip code, 63.8% of respondents thought that this should be categorised as sensitive data under data protection regulations. 31.7% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 5.2% indicated they were not sure/did not know.

4.2.2.14 RO2 EU Survey: Income

Respondents were asked: “How do you believe data revealing your **personal income details** should be classified under data protection laws?”

Figure 4-27 RO2 – Personal Income (EU)



When asked about their data disclosing their personal income details, 87.3% of respondents thought that this should be categorised as sensitive data under data protection regulations. Only 8.9% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 3.8% indicated they were not sure/did not know.

4.2.2.15 Summary: RO2 -New Categories EU Survey

In the table below summarises the responses to the survey questions. The order has been sorted based on questions where the responses have the highest percentages of 'Sensitive data' responses. The categories in **bold** are classified as sensitive data categories under EU data protection regulations.

Table 4-2 Summary of RO2 Responses (EU)

	Data type	Sensitive	Not-sensitive	Not sure
1.	Health or medical data	93.4%	5.2%	1.4%
2.	Physical location or movement	91%	6.1%	2.8%
3.	Personal Income Details	87.3%	8.9%	3.8%
4.	Sexual orientation	84%	13.1%	2.8%
5.	Home address or postcode	75.6%	20.7%	3.8%
6.	Political opinions	67.1%	27.2%	5.6%
7.	Religious and/or philosophical belief(s)	62.4%	31%	6.6%
8.	Current/past employment details	63.8%	31%	5.2%
9.	Trade union membership	50%	39.4%	10.3%
10.	Date of birth	44.9%	47.9%	5.2%
11.	Racial or ethnic origin	41.8%	54%	4.2%

The table above shows that there out of the 11 categories of data presented, there are 5 categories where over 75% of respondents believe these should be classified as sensitive personal data under data protection regulations. Only two of these (i.e. health or medical data and data relating to sexual orientation) are actually categorised as 'sensitive data' under EU regulations.

This means there are three categories of personal data where over 75% of respondents believe should be treated as 'sensitive' data under data protection regulations. The three categories are, Physical location or movement; Personal Income Details; and Home address or postcode. These findings suggest that EU regulators have failed to incorporate respondents views/expectations on what they personal data they deem sensitive into relevant data protection regulations.

We will now go on to examine how US based responded to the same survey.

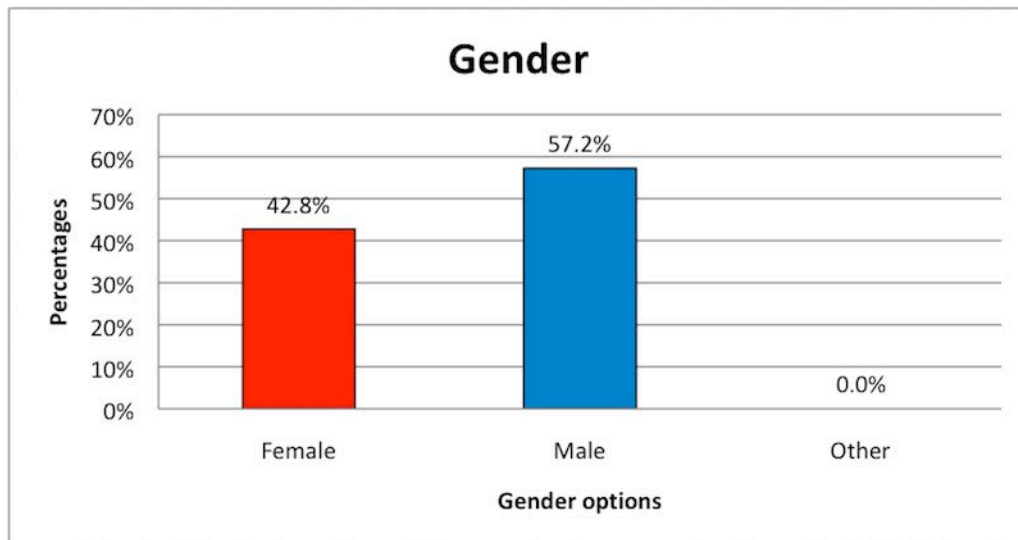
4.2.3 RO2 Sensitive Data: US Findings

The results of the US survey are presented in a series of tables below. MTurk allows requesters to only permit workers registered in the US to answer the survey. There were a total of 201 valid responses in this survey.

4.2.3.1 RO2 US Survey: Gender

Respondents were asked: “How do you believe data revealing your **racial or ethnic origin** should be classified under data protection laws?”

Figure 4-28 RO2 – Gender (US)

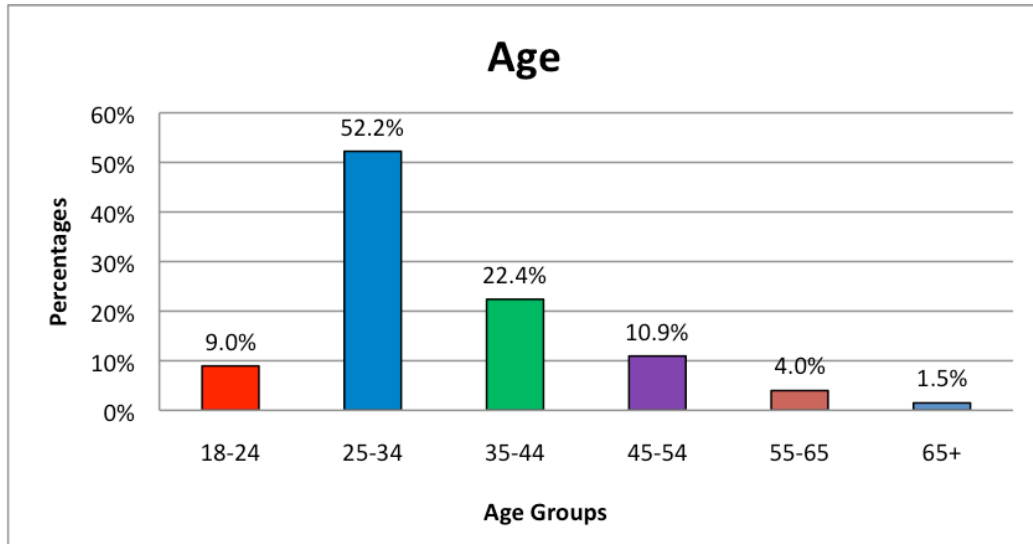


The results show a majority of the survey participants are male and these make up 57.2% of the survey responses. Female respondents made up 42.8% of responses and none of the participants identified as ‘other’.

4.2.3.2 RO2 US Survey: Age

Respondents were asked: “How do you believe data revealing your **racial or ethnic origin** should be classified under data protection laws?”

Figure 4-29 RO2 – Age (US)

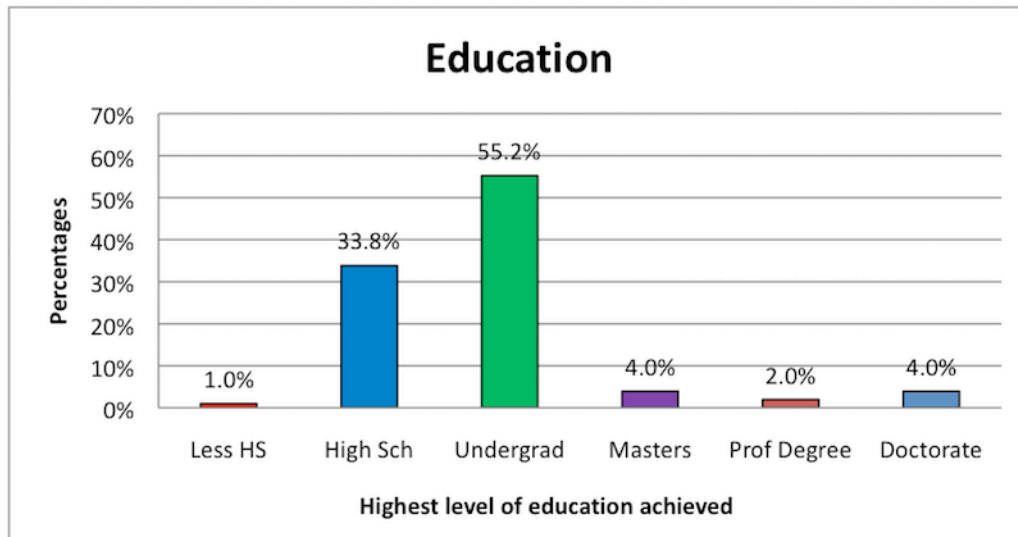


Out of the 201 valid survey responses, the 25-34 year age range were the largest group representing 52.2% of all respondents. Next was the 35-44 age range making up 22.4% of responses. The third largest group were 45-54 year olds at 10.9%. The 18-24 year age range constituted 9% of the sample while 55-65 year range made up just 4% of respondents. The 65+ age range represented 1.5% of the sample.

4.2.3.3 RO2 US Survey: Education

Respondents were asked: “How do you believe data revealing your **racial or ethnic origin** should be classified under data protection laws?”

Figure 4-30 RO2 – Education (US)

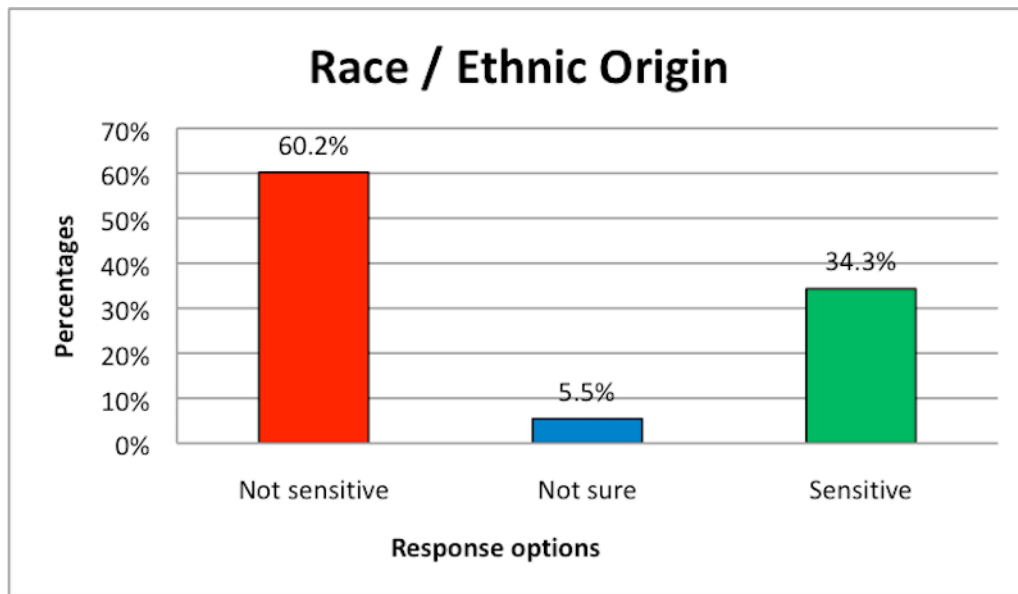


As we can see, 99% of respondents had at least finished high school with 65.2% having completed an undergraduate degree. Over 55% of respondents had achieved an undergraduate degree while a further 10% of participants said they had completed a Masters, Doctorate or professional degree.

4.2.3.4 RO2 US Survey: Racial / Ethnic Origin

Respondents were asked: “How do you believe data revealing your **racial or ethnic origin** should be classified under data protection laws?”

Figure 4-31 RO2 – Race / Ethnic Origin (US)

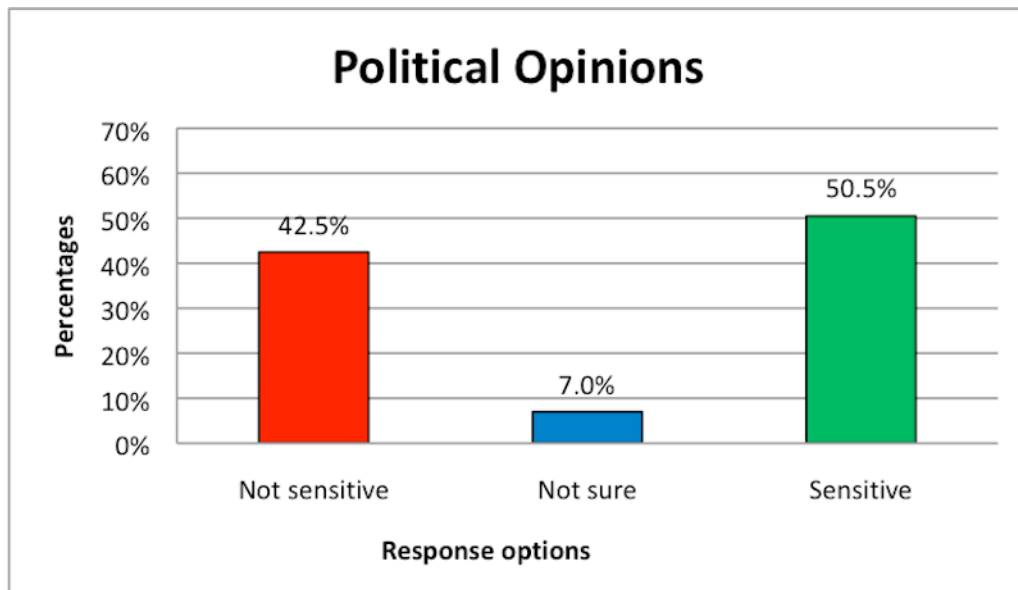


The findings show that out of the 201 valid responses, 60.2% believe details that identify your racial or ethnic origin should not be categorised as sensitive data under current EU data protection regulations. 34.3% of respondents thought it should be classified as sensitive data and 5.5% indicated they were not sure/did not know.

4.2.3.5 RO2 US Survey: Political Opinions

Respondents were asked: “How do you believe data revealing your **political opinion(s)** should be classified under data protection laws?”

Figure 4-32 RO2 – Political Opinions (US)

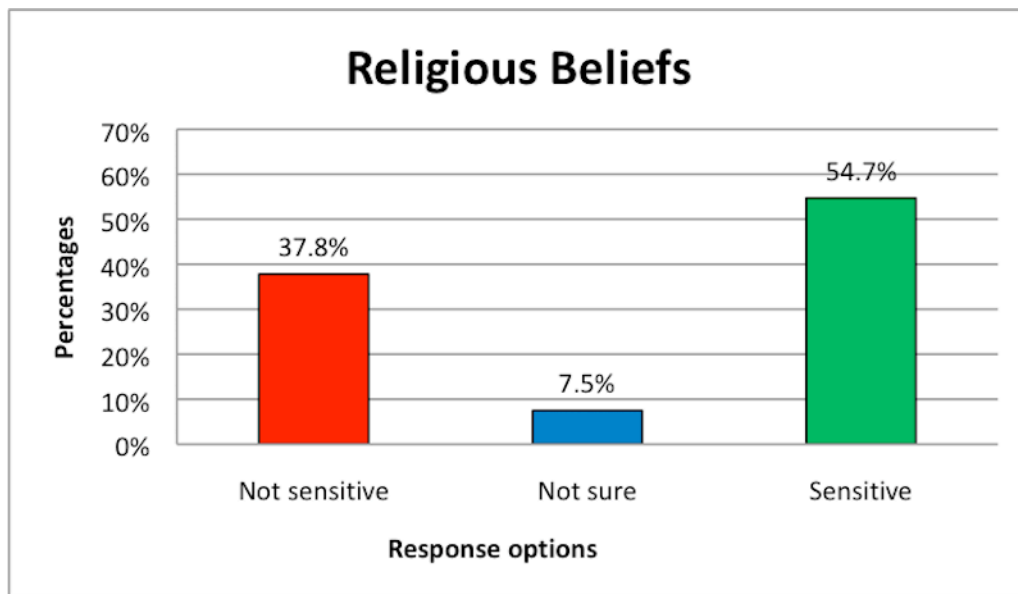


When asked about their political opinions, 50.5% of respondents thought that data revealing political opinions should be categorised as sensitive data under data protection regulations. 42.5% of respondents thought it this category of personal data should not be classified as sensitive data. The remaining 7% indicated they were not sure/did not know.

4.2.3.6 RO2 US Survey: Religious Beliefs

Respondents were asked: “How do you believe data revealing your **religious and/or philosophical belief(s)** should be classified under data protection laws?”

Figure 4-33 RO2 – Religious Beliefs (US)

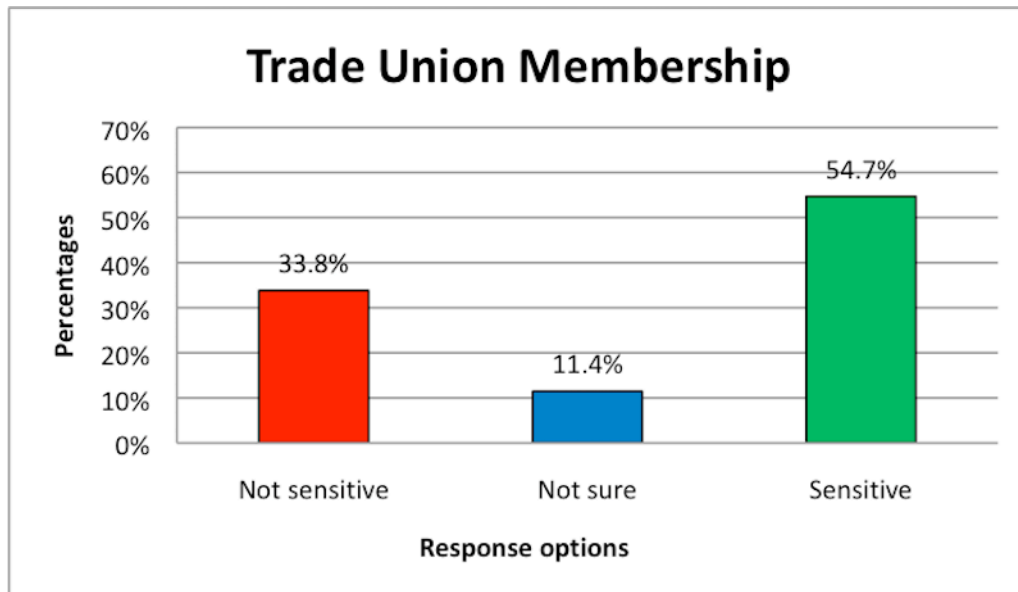


When asked about their religious beliefs, 54.7% of respondents indicated that they thought that this should be categorised as sensitive data under data protection regulations. 37.8% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 7.5% indicated they were not sure/did not know.

4.2.3.7 RO2 US Survey: Trade Union

Respondents were asked: “How do you believe data revealing your **trade union membership** should be classified under data protection laws?”

Figure 4-34 RO2 – Trade Union Membership (US)

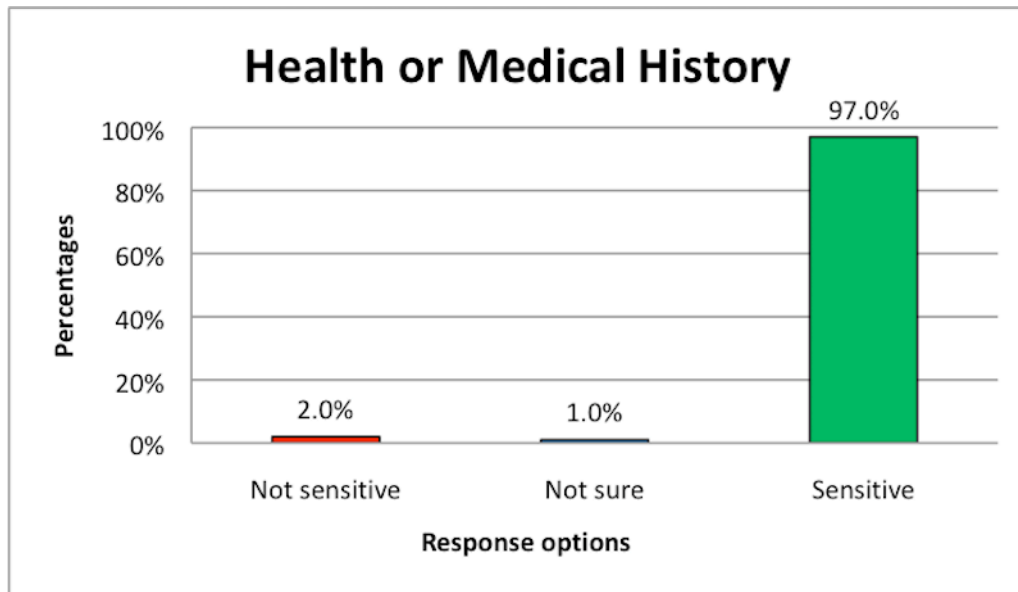


When asked about their trade union membership, 54.7% of respondents thought that this should be categorised as sensitive data under data protection regulations. 33.8% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 11.4% indicated they were not sure/did not know.

4.2.3.8 RO2 US Survey: Health Data

Respondents were asked: “How do you believe data revealing your **health or medical history** should be classified under data protection laws?”

Figure 4-35 RO2 – Health or Medical History (US)

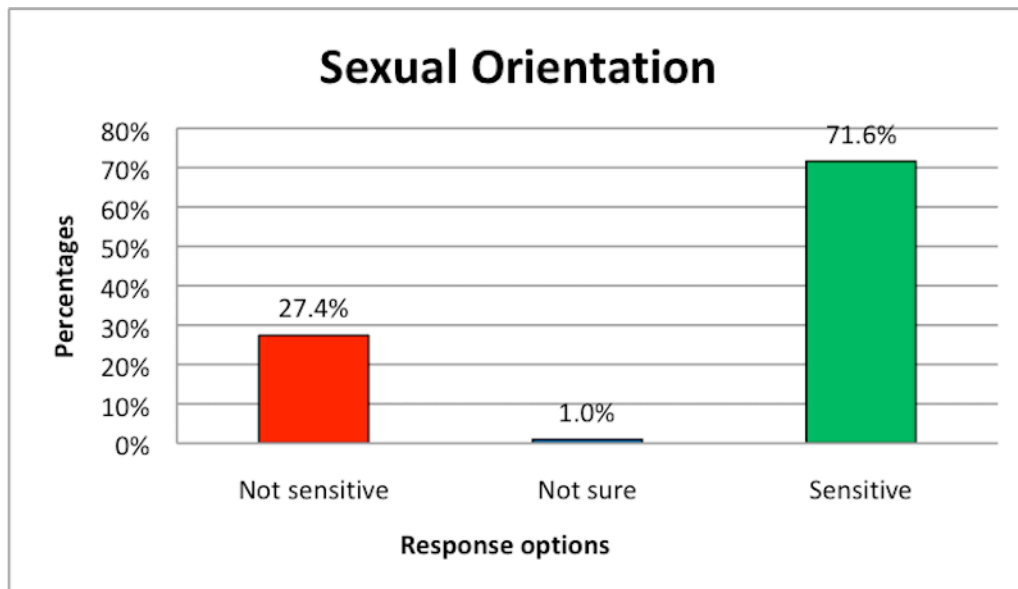


When asked about their data disclosing their health or medical history, 97% of respondents thought that this should be categorised as sensitive data under data protection regulations. Only 2% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 1% indicated they were not sure/did not know.

4.2.3.9 RO2 US Survey: Sexual Orientation

Respondents were asked: “How do you believe data revealing your **sexual orientation** should be classified under data protection laws?”

Figure 4-36 RO2 – Sexual Orientation (US)

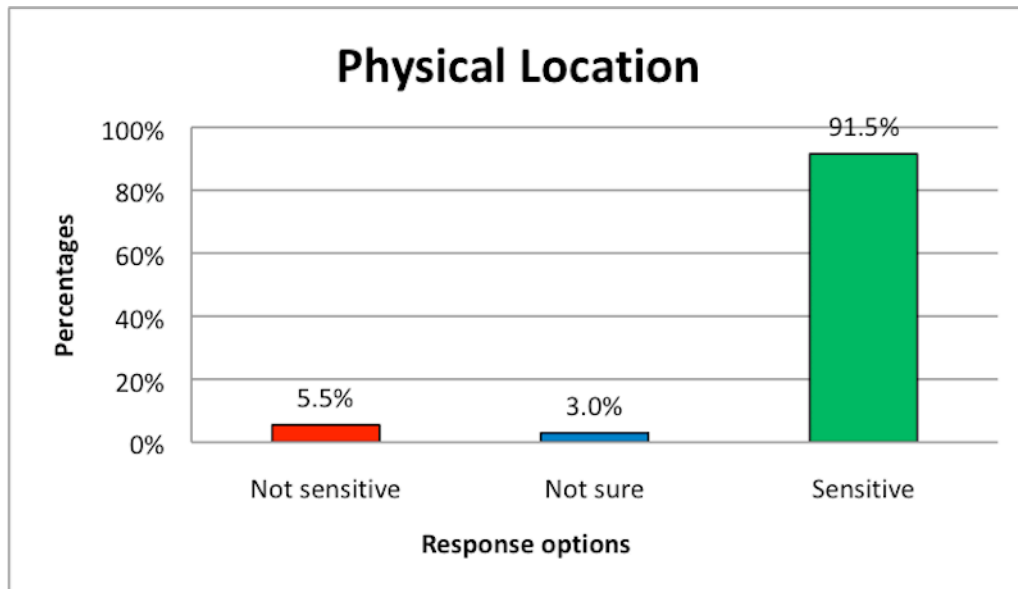


When asked about their data disclosing their sexual orientation, 71.6% of respondents thought that this should be categorised as sensitive data under data protection regulations. 27.4% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 1% indicated they were not sure/did not know.

4.2.3.10 RO2 US Survey: Location

Respondents were asked: “How do you believe data revealing your **physical location or movement** should be classified under data protection laws?”

Figure 4-37 RO2 – Physical Location (US)

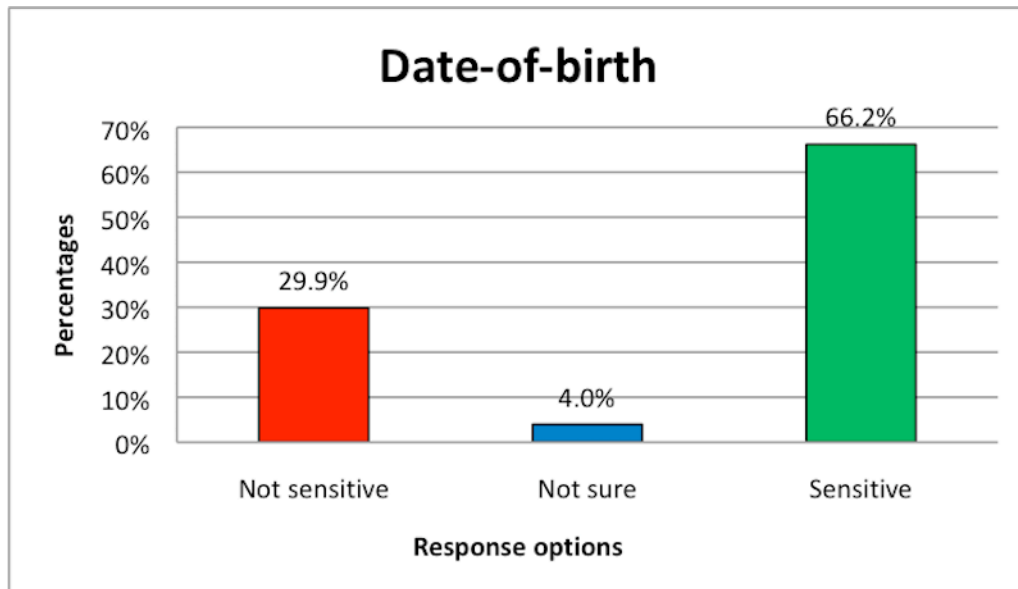


When asked about their data disclosing their physical location or movement, 91.5% of respondents thought that this should be categorised as sensitive data under data protection regulations. Only 5.5% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 3% indicated they were not sure/did not know.

4.2.3.11 RO2 US Survey: Date-of-birth

Respondents were asked: “How do you believe data revealing your **date of birth** should be classified under data protection laws?”

Figure 4-38 RO2 – Date-of-birth (US)

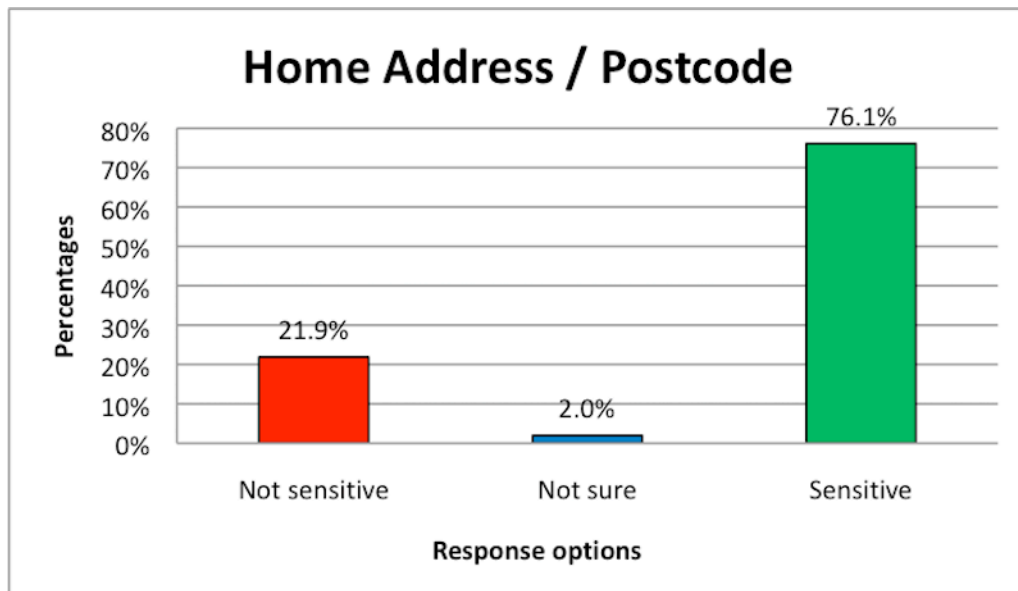


When asked about their data disclosing their date-of-birth the results show that 66.2% of respondent's thought that this should be categorised as sensitive data while 29.9% thought it this category of personal data should not be classified as sensitive data. The remaining 4% indicated they were not sure/did not know.

4.2.3.12 RO2 US Survey: Home Address

Respondents were asked: “How do you believe data revealing your home address postal code should be classified under data protection laws?”

Figure 4-39 RO2 – Home Address (US)

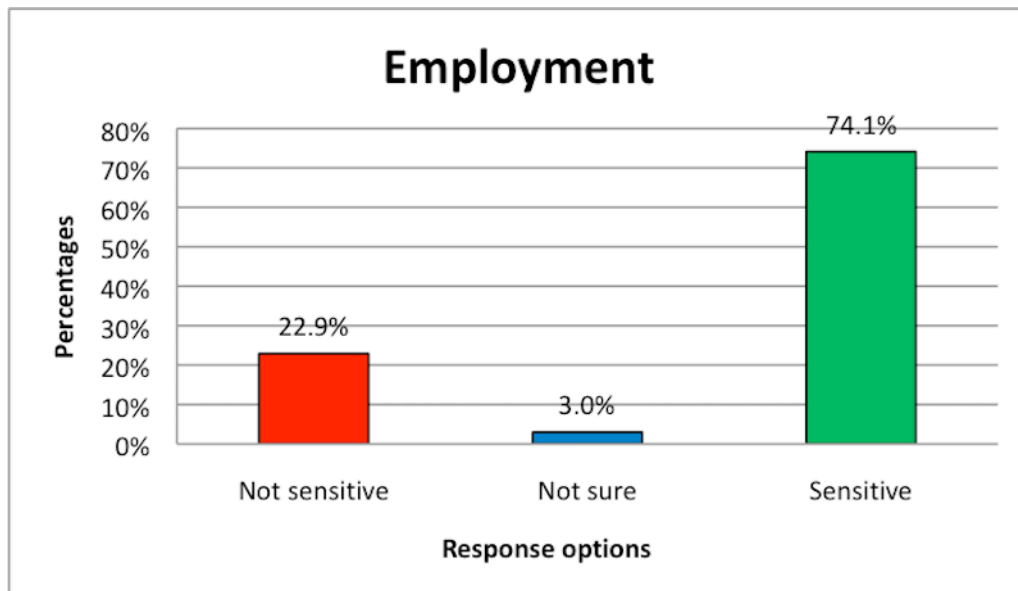


When asked about their data disclosing their how address post/zip code, 76.1% of respondents thought that this should be categorised as sensitive data under data protection regulations. 21.9% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 2% indicated they were not sure/did not know.

4.2.3.13 RO2 US Survey: Employment

Respondents were asked: “How do you believe data revealing your current/past employment details should be classified under data protection laws?”

Figure 4-40 RO2 – Employment (US)

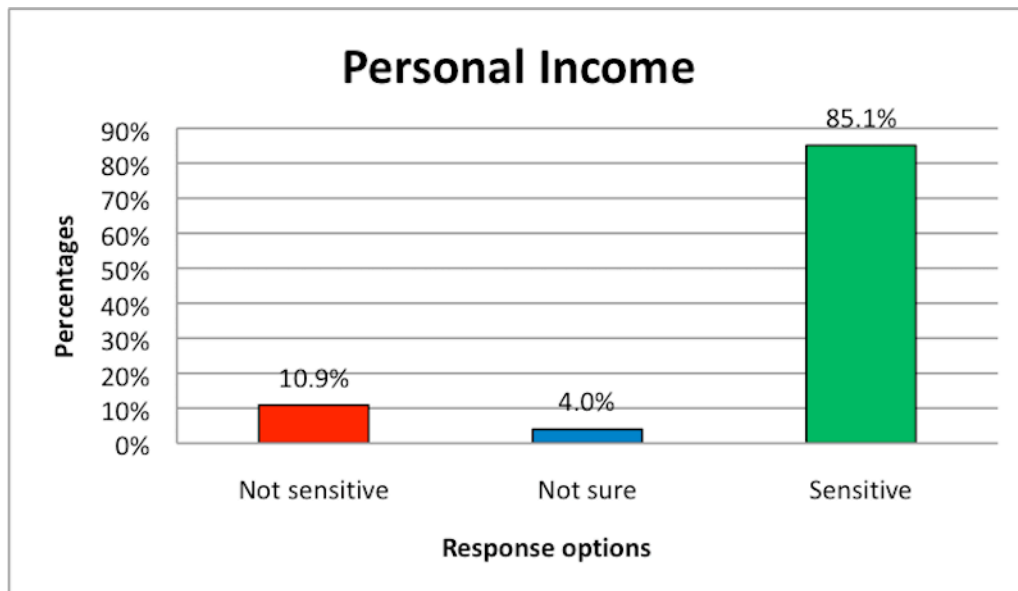


When asked about their data disclosing their how address post/zip code, 74.1% of respondents thought that this should be categorised as sensitive data under data protection regulations. 22.9% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 3% indicated they were not sure/did not know.

4.2.3.14 RO2 US Survey: Income

Respondents were asked: “How do you believe data revealing your **personal income details** should be classified under data protection laws?”

Figure 4-41 RO2 – Personal Income (US)



When asked about their data disclosing their personal income details, 85.1% of respondents thought that this should be categorised as sensitive data under data protection regulations. Only 10.9% of respondents thought it this category of personal data should not be classified as sensitive data and the remaining 4% indicated they were not sure/did not know.

4.2.3.15 Summary: RO2 - New Categories US Data

The findings from the US based survey show there are 5 categories of personal data where more than 74% of respondents thought they should be classified as 'sensitive' personal data. The data categories are, 1) health or medical data; 2) Physical location or movement; 3) personal income details; 4) home address or postcode, and 5) current/past employment details.

Table 4-3 RO2 – Responses Ranked (US)

	Data type	Sensitive	Not Sensitive	Not sure
1.	Health or medical data	97%	2%	1%
2.	Physical location or movement	91.5%	5.5%	3%
3.	Personal Income Details	85.1%	10.9%	4%
4.	Home address or postcode	76.1%	21.9%	2%
5.	Current/past employment details	74.2%	22.9%	3%
6.	Sexual orientation	71.6%	27.3%	1%
7.	Date of birth	66.2%	29.9%	4%
8.	Religious and/or philosophical belief(s)	54.7%	37.8%	7.5%
9.	Trade union membership	54.7%	33.8%	11.4%
10.	Political opinions	50.2%	43.3%	7%
11.	Racial or ethnic origin	34.3%	60.2%	5.5%

US data protection regulations do not include a specific 'sensitive' data provision as per EU data protection provisions. However, it is worth noting that out of the top five 'sensitive' data categories from this US base survey, only 'Health or medical data' would be categorised as 'sensitive' data in an EU jurisdiction.

The following sections will compare the EU and US based survey findings.

4.2.4 RO2 EU v US Survey: Analysis & Discussion

As discussed in the literature review, EU and US data protection regulation regimes emerged from very different founding principles. The centralized, standard setting EU approach contrasts with the narrow and fragmented US regulatory environment considered the right to privacy as an individual right (Shaffer 2000).

However, when we compare the results of the survey responses we find that both EU and US respondents identified the same ‘top’ three categories of data with the highest percentages indicating they believe these should be categorised as ‘sensitive personal data’. These are: 1) Health or medical data, 2) Physical location or movement and 3) Personal income details. These results give us some interesting insights.

Table 4-4 Differences between US and EU Survey Responses

Data type	EU Sensitive	US Sensitive	Difference (EU – US)
Health or medical data	93.4%	97%	3.6%
Physical location or movement	91%	91.5%	0.5%
Personal income details	87.3%	85.1%	2.2%
Home address or postcode	75.6%	76.1%	0.5%
Current/past employment details	63.8%	74.2%	10.4%
Sexual orientation	84%	71.6%	12.4%
Date of birth	44.9%	66.2%	21.3
Religious and/or philosophical belief(s)	62.4%	54.7%	7.7%
Trade union membership	50%	54.7%	4.7%
Political opinions	67.1%	50.2%	16.9
Racial or ethnic origin	41.8%	34.3%	7.5%

The findings suggest that the categories sensitive data as defined in the GDPR while clearly relevant, can also be considered somewhere limited out of sync with our respondents’ views. This suggests that regulators have failed to keep regulations up to date with what data subjects believe should be categorised as ‘sensitive’ personal data. This is evidenced by the fact that over 90% of EU and

US respondents indicated that data revealing 'physical location or movement' (i.e. location/GPS data) should be categorised as 'sensitive personal information'. Similarly, over 85% of EU and US respondents believe data revealing 'personal income details' should be classified as sensitive data under data protection laws. Neither of these categories is classified as such in the GDPR.

The findings also show that the list of categories defined as 'sensitive personal data' in the GDPR tend to be 'lower' down the rankings in terms of the percentages of respondents who think they should be categorised as 'sensitive personal data'. Examples of this include respondents views in terms of the political opinions and race / ethnic origin categories etc. However, it is important to note that the only category of where more than 50% of respondents selected the 'not sensitive data' response was 'racial or ethnic origin' (EU: 54%, US: 60.2%). These findings are consistent with research from Evens and Van Damme (2016) where consumers are happy to disclose basic demographic personal data but are more concerned with regards to sharing data on income and related financial details.

The category with the largest percentage difference between EU and US respondents was 'Political opinions' where EU respondents it ranked 6th highest percentage of 'sensitive personal data' responses at 67.1%. This contrasts with the US findings where 'political opinions' was ranked 10th with 50.2% of respondents believing this should be classified as sensitive personal data. Given the recent media coverage in relation to how personal data from Facebook and other social media platforms may have been used to influence EU and US elections, it will be interesting to see if these survey findings are consistent over time.

4.2.5 RO2: Summary

The findings from the surveys conducted in relation to RO2 reveal some interesting views on what categories of personal data EU and US based respondents consider 'sensitive' personal data. The findings show that there are

clearly some categories of personal data where there is a difference of over 10% in terms of EU and US respondent's views on what should be classified as 'sensitive' personal data. These categories include 'Current/past employment details', 'Sexual orientation' 'Date of birth' and 'Political opinions'.

The findings also show that there are 7 categories of personal data where there was a less than 10% difference between EU and US respondents' views. However, the most striking is how both survey findings reveal the same top four categories of personal data with the highest percentages of respondents indicating that the category should be classified as 'sensitive' personal data. The difference between EU and US findings are less than 4% for each of these four highest ranked 'sensitive' data categories. They are, 'Health or medical data' (3.6%), 'Physical location or movement' (0.5%), 'Personal income details' (2.2%) and 'Home address or postcode' (0.5%).

The consistency we see in these highest ranking 'sensitive' data categories is useful as we move onto RO3 where we conduct randomised experiments and test what impact a privacy seal has on personal information disclosure. Specifically we will examine what impact, if any, the presence of a privacy seal has on categories of personal data that we know from RO2 findings that both EU and US respondents believe are 'sensitive' personal data. The findings in relation to RO2 help to inform how we set up the RO3 experiments.

4.3 Research Objective 3

Conduct experiments to test if use of 'privacy seals' effect data subjects personal information disclosure.

The sections below briefly summarise the key literature relevant to RO3 stated.

To recap, the GDPR set out provisions for,

"the establishment of certification mechanisms and data protection seals and marks should be encouraged, allowing data subjects to quickly assess the level of data protection of relevant products and services" (Regulation 2016, p.19).

Certification mechanisms, seals and marks also have the potential to meet US regulators stated aims to have,

"easily understandable and accessible information about privacy and security practices. At times and in places that are most useful to enabling consumers to gain an meaningful understanding of privacy risks and the ability to exercise Individual Control" (White House 2012, p.108).

Existing empirical literature specifically examining the use of privacy seals is limited. The following sections summarise the research that does exist in relation to privacy seals as well as relevant related research in this area is.

All relevant empirical findings from the experiments conducted in this research are considered in the context of potential implications for the key stakeholders i.e. data collectors, data subjects and regulators.

4.3.1 Addressing Information Asymmetry

Information asymmetry is a key area of concern in the relationship between those who used new media service users (i.e. data subjects) and new media service providers (i.e. data controllers). Traditionally data controllers have used privacy policies to state how, why and to whom etc they share personal

information they collect from data subjects. However, a number of research studies show that the length and complexity of these policies mean data subjects rarely read them (Vila et al 2004; Jensen et al 2005; Reidenberg, 2015). Information sharing has the potential to be used to discriminate against the individual, or may even result in identity theft. Research findings from Miyazaki and Krishnamurthy (2003) suggests that data subjects can mistakenly interpret the presence of a privacy policy (or link to it) as an assurance that their data is protected.

4.3.2 Iconography & New Media

The use of privacy seals (or similar) can be considered to fall under the broad category of 'iconography' in academic literature. As examined in the literature review, iconography has been used in a range of online and offline settings in an attempt to reduce information asymmetry i.e. a position where one party in a trade is in possession of information that another party(s) in the same trade is not (Edwards and Able, 2014; Bertoldi and Atanasiu, 2007; Grunert and Wills, 2007).

The standardized energy labels for domestic white goods provides a useful example of how relatively complex information (i.e. relating to energy efficiency, and consumption and other key product information) can be visually conveyed to purchasers. The success of the use of standardized energy labels in reducing information asymmetry in the domestic appliance market appears may have influenced the decision to include provision for a privacy seal accreditation system in the GDPR although regulator do not explicitly state this. At a macro level the scheme achieved its goal of reducing energy consumption while also enjoying the micro success of changing user behaviour, with purchasers clearly preferring to purchase products with lower energy consumption ratings.

4.3.3 Privacy Seals: RO3 Research Findings

There is relatively little literature that specifically examines the impact privacy seals have in terms of how they impact on personal information disclosure. The studies that do exist (Rifon et al 2005; John et al 2010; Brandimarte et al 2013; Carolan & Castillo-Mayen 2015; Tsai et al 2011) have produced inconclusive and at times counter-intuitive findings that are relevant to the experiments conducted in relation to RO3. Key research outputs are summarised in the following sections.

4.3.3.1 Issues with miss-interpretation

Specifically examining the effect of privacy seals in online environments, experimental research from Rifon (2005) suggests that participants recognise the assurances promised by the presence of a privacy seal. However, the findings show that participants miss-interpreted the presence of a privacy seals.

Privacy seals were also found to have also increased participants belief that the website displaying the seal would inform them of its data collection and sharing practices. The study suggests that,

“An evaluation of actual disclosure response to privacy seals is a challenge to academic researchers but appears to be a next step in evaluating the effectiveness of seals” (Rifon 2005, p.359).

This thesis examines the very subject of actual disclosure when a privacy seal is displayed.

4.3.3.2 Contextual Clues

The presence of a privacy seal may provide act as a ‘contextual clue’ to data subjects. Research by John et al (2010) confirms that contextual clues have an impact on information disclosure. Their experimental research findings demonstrated the somewhat counter-intuitive impact of such ‘clues’. The research output shows that respondents are more willing to disclose personal

information in the condition where the information is requested on an unprofessionally designed web page. In contrast to this, experiment participants were less willing to disclose personal information in a condition where the same information was requested from a more professionally designed web page.

In the cases examined, contextual clues in the form of a more professionally designed interface and/or a warning about potential privacy concerning practices (i.e. phishing) appears to have alerted privacy concerns for data subjects resulting in less personal information disclosure. As a privacy seal has the potential to act as a contextual clue, these findings are important.

4.3.3.3 The Peltzman effect

The 'Peltzman effect' is an observed condition where the wearing a car seat belt makes the driver 'feel' safer. This can result in them driving more recklessly and therefore increase their likelihood of being in an accident. Experimental research from Brandimarte et al (2012) sees findings analogous to the Peltzman effect. Their research findings demonstrate that individuals fail to distinguish between *control over the release* of personal information and *control over access and usage* of personal information. The findings suggest that technological features designed to make individuals feel more in control over the release of their personal information, may have the unintended result of inducing riskier information disclosure behaviour from data subjects. These findings potentially have important implications for this research.

4.3.3.4 Unintended Consequences

Research findings from Carolan & Castillo-Mayen (2015) looked at the use of cookies and consent regulations as set out by EU regulators under the 2002 e-Privacy Directive. Although this research did not specifically look at the use of privacy seals, their findings demonstrate how mechanisms designed to empower data subjects with greater control may in fact result in information disclosure counter to their intentions. The authors believe their findings cast

doubts on the efficacy of legal strategies that assume the provision of clear information (i.e. in this case about cookies) will influence user behaviour in any sort of predictable manner.

4.3.3.5 Privacy Ratings

Research from Tsai et al (2011) looked at the impact privacy ratings had on user purchased behavior. In their experiments, they modified search engine results to display an icon indicating a fictitious privacy rating for merchants listed in search results. Their findings showed that participants were more likely to purchase from websites with high or medium privacy rating icon was displayed. They were also prepared to pay a premium for products where the merchant had a higher privacy rating. These results potentially have significant implications for data collects, data subjects and regulators.

4.3.4 Summary

The GDPR includes a provision for the establishment of a privacy seal accreditation system. Such schemes are a clear attempt by regulators to address issues relating to information asymmetry. In the domestic appliance market, EU regulators successfully introduction standardized labels/icons to reduce similar information asymmetry issues that existed in this market. Regulators clearly see the potential for a privacy seals to help in relation to data protection.

There is a relatively limited body of empirical data that looks at the effects of privacy seals on information disclosure in online environments. We also looked at other relevant research areas where the use of privacy seals has the potential to as a contextual clue, a de facto legal notice, a privacy rating metric as well it potentially inducing a Peltzman type effect.

The following sections will examine the findings of the experiments carried out in this research study. Specifically, we will examine the impact the presence of

a privacy seal has on personal information disclosure. Where relevant, we will discuss how these findings impact these findings have on existing privacy seal (and related) empirical studies.

4.3.5 RO3: Findings, Analysis & Discussion

The tables and sections below present the key findings from the experiments conducted in relation to RO3. The findings are discussed in the context of how it extends (or otherwise) existing research findings as well as discussing the potential implications for data subjects, data controllers and regulators.

Participants in each of the experiments were randomly assigned to a control or treatment group by the Qualtrics software application. Those assigned to the Control group were asked to complete the survey questions (as detailed in the methodology chapter). Participants assigned to the Treatment group asked to answer exactly the same survey questions. The only difference between the Control and Treatment groups was the presence of a Privacy Seal in the header and footer of each page that those in the Treatment group. A section in the survey asks participants to '*Please tell us a little bit about yourself*' and here 7 x categories of personal data are requested.

Using the findings from RO2 (presented earlier), we include the 5 categories that the highest percentage of respondents indicated was 'sensitive' personal data. For all questions in this section, a '*would rather not say*' option is available from the drop-down list of available answers. To test for statistically significant differences between the groups we use the Fisher Exact test and the accompanying p-value is reported. As is common with most statistical methods, we would expect to see a p-value lower than 0.05% if we are to accept the differences in responses between the two groups are statistically significant.

4.3.5.1 RO3: EU Experiment – Text only v Privacy Seal

The table below shows the findings from the EU based experiment.

Figure 4-42 RO3 – EU Data: Fisher Exact Test Results

EU Experiment Summary Data: Text v Privacy Seal	
Category	P-values
Gender	p-value 1.0
Age	p-value 1.0
Education	p-value 1.0
Personal income	p-value 0.6201
Sexual Orientation	p-value 1.0
Personal Health	p-value 0.6854
Location	p-value 1.0

Note: Control Group = Text only and Treatment Group = Privacy Seal

Total number of participants: n = 152 [Control Group: 71 | Treatment: 81]

Figure 4-43 Privacy Seal Image Used In EU Experiment



As shown above, none of the p-values are below the 0.05 value that we would expect to see if there was a statistically significant difference in terms of respondent's decisions to disclose personal information between control and treatment groups.

Although the experiments did reveal differences in disclosure for the 'Personal Income' and 'Personal Health' categories, these differences were not considered statistically significant. The findings show that presence of the privacy seal did not have a statistically significant impact on the experiment participant's information disclosure in any of the 5 sensitive data categories.

4.3.5.2 RO3: US Experiment – Text only v Privacy Seal

The table below shows the findings from the US based experiment. The survey questions were the same but the privacy seal image used was different i.e. used the ePrivacy icon as this did not have the word ‘Euro’ in it and would therefore be more relevant to a US audience.

Figure 4-44 RO3 – US Data: Fisher Exact Test Results

US Experiment Summary Data: Text v Privacy Seal	
Category	P-values
Gender	p-value 1.0
Age	p-value 1.0
Education	p-value 1.0
Personal income	p-value 0.6201
Sexual Orientation	p-value 1.0
Personal Health	p-value 0.2452
Location	p-value 0.4968

Note: Control Group = Text only and Treatment Group = Privacy Seal

Total number of participants: n= 152 [Control Group: 77 | Treatment: 75]

Figure 4-45 Privacy Seal Image Used In Us Experiment



The US experiment findings were similar to what we found in the EU. None of the Fisher Exact Test p-values can be considered statistically significant (i.e. > 0.05%). The ‘Personal Income’, ‘Personal Health’ and ‘Location’ categories did show differences between control and treatment groups albeit the differences were not significant to say the presence of a Privacy Seal caused differences in personal information disclosure.

4.3.6 RO3 - Key Findings: Analysis & Discussion

The results of the EU and US based experiments extend existing research findings in this area in a number of ways.

Consistent with the findings from Rifon (2005) the presence of a privacy seal in our experiments did not influence personal information disclosure either positively or negatively. However, it is important to note that the research conducted in this thesis did not explicitly measure 'trust' so we cannot evaluate whether the presence of the seal engendered more trust as it did with La Rose. It is also important to note that there are differences between how the La Rose experiments and the experiments conducted in this research. For example, their privacy seal indicated compliance with their stated privacy policy whereas the privacy seals used in the experiments conducted here indicate compliance with data protection regulations. However, on the core issue of privacy seals not impacting the amount of personal information disclosed the findings are indeed consistent.

Existing research has also examined the role contextual clues play in personal information disclosure. Findings from John et al (2010) showed that the display of text warnings about potential privacy violations triggered privacy concerns and reduced information disclosure. In the experiments conducted for this thesis, the presence of a contextual clue in the form of a privacy seal did not lead to any statistically significant difference in personal information disclosure. This was consistent in both amongst EU or US respondents. One might expect, and indeed we observed some respondents in both the control and treatment groups would choose not to disclose this type of personal information. However, the differences between the control and treatment groups were shown by the Fisher Exact test not to be statistically significant. Therefore, we can say that the presence of the privacy seal did not have any positive or negative 'contextual clue' causal impact on personal information disclosure.

This has potentially very important implications for data controllers. The findings here offer no evidence that the display of privacy seals 'trigger privacy concerns' that in turn lead to reduction in personal information disclosure. This is an important finding for new media firms. Displaying a privacy seal that

reduces personal information disclosure would act as a disincentive for firms to participate in such an accreditation scheme and these findings have positive impacts for data subjects, data controllers and regulators alike.

As discussed earlier, the so-called 'Peltzman effect' observed that seat belts make drivers feel safer, and the in turn drive faster and are therefore more likely to be involved in an accident (Brandimarte et al 2013). As applied to personal information disclosure, the display of a privacy seal (or other privacy enhancing technologies) may result in more 'reckless' personal information disclosure exposing individuals to greater risk of identity theft, fraud etc. The results of the experiments conducted do not show any evidence of the presence of any such 'Peltzman effect'. However, with at times very high levels of personal information disclosure in the control groups, it was not possible to estimate significantly greater disclosure in treatment groups so this finding should perhaps be observed with this limitation in mind.

Similarly, there is no evidence of the 'unintended consequences' observed by Carolan & Castillo-Mayen (2015) where mechanisms designed to empower data subjects with more control is not observable in any sort of predictable manner. The evidence from the experiments conducted here suggests a privacy policy that independently verifies compliance with data protection regulations has no statistically significant (i.e. higher or lower level of disclosure) impact on personal information disclosure.

When taken in conjunction with the findings in relation to RO2, these experiment findings also add further evidence to the existence of a 'privacy paradox'. Although both US and EU respondents identified categories of data that they identified as 'sensitive personal data', the experiments showed that many were willing to disclose this personal information for a relatively small financial reward i.e. \$0.80. This can be seen as an example of where stated privacy concerns do not match actual behaviours (Taddicken 2014).

The findings make it difficult to draw any conclusions in relation to regulatory key concern in addressing information asymmetry.

5 Chapter 5: Conclusions

This chapter presents the conclusions of the key findings in this thesis. These conclusions are considered in the context of implications for new media firms as well as for policy makers and regulators.

Lucy Küng (2008) reminds us of the important wider role new media firms play in society and that they must behave in the public interest and not focus solely on profits and shareholder value,

“media is a cultural force: it shapes society and its messages are fundamental to democracy” (Küng 2008, p.11).

Leading media management academic Professor Robert Picard (2002) is clear on the role of policy makers and regulator.

“Regulations exist to facilitate communications as well as to promote or constrain certain types of communications activities” (Picard 2002, p70).

Data protection regulators play an important role balancing the needs of the individual with the wider information needs of the society (Cate 1999). EU and US regulators have highlighted issues with how technology progress is impacting on personal privacy. In the EU, the GDPR seeks to ensure that technological advances and related services, pioneered by the new media sector, operate in the public interest. Regulators have highlighted the unequal relationship that exists between new media service providers (data collectors) and users of these media services (data subjects). This information asymmetry means that media consumers are not always clear what personal data media firms are collecting from them and how it is used/shared etc. Regulators identify control and transparency as a key to reducing the effects of information transparency, thus ensuring public interest is protected. To that extent, many aspects of the GDPR can be thought of as a form of behavioural regulation that media firms have contended with for decades. This type of behavioural regulation,

“prohibits media firms from engaging in certain practices or requires them to engage in specific practices” (Picard 2002, p.71).

EU regulators, through the GDPR, set out to strengthen protections for personal data in light of advances in technology. While the GDPR is a comprehensive and wide-ranging regulation, this thesis focusses on the regulations provision for the establishment of a privacy seal accreditation scheme. While participation in privacy seal accreditation schemes is not a legal requirement for data controllers, regulators do see it as a tool to give data subjects (i.e. new media users/consumers) enhanced control and transparency over their personal data.

There is no empirical evidence in relation to the impact the specific type of privacy seal envisaged by the GDPR would have on new media consumers or service providers. This thesis addresses this research gap within media management academic research. This is done in light of the lack of consensus on common standards to protect personal information (Jayakar 2018).

This discussion will also identify some of the limitations of this research as well as identifying avenues for future media management research.

The overarching aim of this thesis examines how ‘new’ provisions in the GDPR impact personal information disclosure in online environments. In order to achieve this, online surveys and experiments were conducted to answer three set research objectives. The following sections we present the conclusions that can be drawn in relation to each research objective.

5.1 Research Objective 1: Conclusion

The conclusion drawn from the survey findings in relation to RO1 is that high percentages of the EU based survey respondents are successfully able to identify the categories of personal data that are categorised as ‘sensitive data’ under EU data protection regulations. The findings are also important from an internal validity perspective. Given the methodological approach adopted in this thesis, it was important to establish that ‘sensitive personal data’ was a concept

that was broadly understood by survey participants before moving on to the other research objectives. The findings clearly show that large numbers of survey respondents are able to identify categories of personal data identified as 'sensitive data' under EU data protection regulations.

The findings here provide important insights for new media firms. The EU respondents surveyed do understand that some categories of data are classified as 'sensitive' data. As public awareness continues to grow around how advertisers use personal information to target for advertising, including political advertising, media firms need to be aware that new media consumers can differentiate between sensitive and non-sensitive data (as per the relevant regulations). In the context of this thesis, the results represent an important building block for issues addressed in RO2 and RO3 i.e. that sensitive personal data categories is a concept understood by respondents.

5.2 Research Objective 2: Conclusion

RO2 set out to establish what categories of personal data respondents believe *should be* classified as sensitive data under data protection rules. As the survey questions are not specific to one regulatory jurisdiction, both EU and US respondents participated in the survey. The surveys were distributed to these two separate groups.

Despite the differing data protection regulatory provisions in the two jurisdictions, the survey results showed very similar results. Among both EU and US respondents', the same three categories of personal data attracted the highest percentages of respondents selecting as 'sensitive personal data' i.e. 1) Health or medical data, 2) Physical location or movement and 3) Personal income details.

The EU survey results show that out of the top five categories of personal data that respondents consider sensitive personal data, only two of these are actually categorised as 'sensitive data' under existing data protection regulations. These are 'Health or Medical' data and data relating to 'Sexual

orientation'. For EU data subjects, data protection regulations in relation to sensitive data categories clearly do not match their expectations. This is evidenced by the fact that 91% of EU respondents believe data revealing 'Physical location or movement' and 87.3% believe that 'Personal income details' should be classified as sensitive personal data. The figures for US respondents were similarly high at 91.5% for 'Physical location or movement' and 85.1% for Personal Income details.

A key conclusion drawn from the findings in relation to RO2 is that EU regulators have failed to update the sensitive data categories in the GDPR in line with data subjects' expectations. This is despite a number of discussion documents and white papers in both the EU and US did consider/propose expanding the list of categories of 'sensitive' data, only very minor amendments were made to the 95/46 directive. Barring the inclusion of biometric and genetic data, the list of categories of 'sensitive' data in the GDPR remain the same as those defined in the 1953 convention on human rights. This seems remarkable given the rate at which technology has developed over the last 65 years, the list of sensitive data categories barely expanded.

From a regulator perspective, the GDPR represents the first major overhaul of data protection in the EU for over 20 years. These survey findings can be viewed as a missed opportunity for regulators to meet the expectations of data subjects with regards to sensitive personal data protections. The survey findings clearly show that the limited set of categories of data that regulators afford 'sensitive personal data' status to, are not in line with data subjects' expectations. As such, regulators need to expand the categories of sensitive data.

For data controllers like new media firms, these findings have a number of potential implications. Obtaining data subjects GPS location and/or IP address allows advertisers to target data subjects based on their locations in order to offer more personalized features/content/advertising. Many applications and services offered by new media firms request information like GPS location from users devices to be switched 'on' in order for users to have access to all of the applications functionality. Similarly, many applications routinely collect the IP

address of the user, which in turn allows for location identification. It is clear from this survey that the vast majority of respondents think this type of data should be considered sensitive data.

However, the findings can also be viewed as an opportunity for new media service providers to listen to their audiences and meet their expectations in relation to how their personal data is managed, rather than purely adhering to regulations.

Almost all neoliberal macroeconomic growth models identify 'technological growth' as the key determinant in delivering long term increases in standards of living. The GDPR can be viewed as an example of where regulators having had to consider the conflicting requirement of the individuals fundamental right to privacy versus the prevention of obstacles to economic growth (Maxeiner 1995) and the lobbying and arguments of the new media firms and the sectors role in delivering economic growth won out.

These findings also demonstrate that media audiences' beliefs about sensitive data are not distinguished by jurisdiction and this suggests that media firms may benefit from meeting audience preferences rather than merely adhering to the regulatory rules for specific environments.

Furthermore, these findings highlight a potential opportunity to capitalise/exploit the 'gap' that exists between data protection regulations and data subjects' expectations. Media firms could choose to differentiate themselves in the market and compete with rivals by explicitly stating that they will give additional assurances/protections to specific data categories, beyond what is required by GDPR (or other relevant regulation).

In 2018 it emerged that a 'personality quiz' distributed on Facebook was used to harvest the personal details of 87 million users. The harvested data was then used to target users with customised political messages (O'Kane 2018). While there is no suggestion that anyone broke the law, the subsequent 'Delete Facebook' campaign highlights that for new media firms, even the perceived misuse of users personal data can impact negatively on the firm. Facebook

founder and chief executive Mark Zuckerberg sees it as a breakdown of trust. In an interview with Time magazine he said,

“I think it’s a clear signal that this is a major trust issue for people, and I understand that. And whether people delete their app over it or just don’t feel good about using Facebook, that’s a big issue that I think we have a responsibility to rectify.” (Lang 2018)

Numerous high profile personal data breaches at major new media providers further highlights the risks to data subjects. In September 2018, Facebook announced that a ‘targeted attack’ had successfully accessed data from approx. 30 million user accounts. Personal data including name, email address, phone number, date of birth and recent locations were taken from users (including Facebook founder and CEO Mark Zuckerberg as well as COO Sheryl Sandberg accounts’). Interestingly, Facebook sought to reassure its users that more sensitive data like sexual orientation were not accessed in the hack as this data was kept more security. However, data was stolen on the most recent searches entered in users Facebook search bar. This has led to concerns that Facebook users were put at risk,

“hackers could potentially target victims with blackmail scams threatening to reveal this info to the world, especially since the hack included user contact info, including phone numbers and email addresses” (Constine 2018).

If, as per our survey findings, more categories of personal data were categorised as ‘sensitive’, then new media firms would be forced to keep them more secure and successful personal data hacks as described would expose data subjects to less risk. There may be an opportunity for new media firms to go beyond purely what regulators require and offer additional assurances to keep categories of personal data like location data, home address details etc (i.e. the categories of data they think *should* be sensitive). This type of policy could be used to rebuild trust between data controller and data subjects.

Continued and sustainable growth in the new media sector is reliant on the confidence of data subjects and failure to address data subjects concerns may

inhibit this growth in the medium to long term. As more major personal data theft/breaches come to light, combined with how personal data can be used to influence national elections etc, there is an opportunity for media firms to offer enhanced privacy protection to compete for users.

We know that data subjects can differentiate between different categories of data and across US and EU survey respondents there is a high degree of consistency in terms of what categories the majority of them consider most/least sensitive. There is an opportunity for firms to compete for consumers by offering greater privacy assurances.

5.3 Research Objective 3: Conclusions

The GDPR includes a specific provision for the establishing a privacy seal accreditation scheme. This is a new provision that did not exist in the 95/46 directive. RO3 set out to examine what impact the presence of a privacy seal has on personal information disclosure.

The conclusion from the experiments conducted is that respondents in the treatment condition where a privacy seal is displayed do not disclose statistically significantly different levels of personal information than those who do not see the privacy seal. These findings are in contrast with existing empirical research in this area which showed that privacy seals (and/or related icons etc) can trigger privacy concerns in data subjects, resulting in lower personal information disclosure. For many new media firms, any mechanism that lowers personal information disclosure has the potential to reduce the opportunities for personalisation and related the associated significant advertising income. Any measure that negatively impacted on revenues would represent a significant disincentive for firms to engage with such schemes. Why would any firm incur the costs associated with participating in an independently accredited privacy seal scheme if it effectively put you at the competitive disadvantage?

The findings clearly show that the presence of a privacy seal does not result (in statistically significantly way) in lower levels of personal information disclosure compare to those who do not see a privacy seal. An obvious question here is to ask why this might be the case when existing research showed it resulting in a lower personal information disclosure effect? It could perhaps be because users awareness of online privacy issues, as highlighted by significant data breaches in recent years, has led to more familiarity with privacy enhancing technologies. Another potential explanation could be that experiment participants simply did not see/notice the presence of the privacy seal.

5.3.1 Implications for Media Firms

The findings do offer a potential opportunity for new media firms to use display privacy seals and differentiate themselves in their market. In recent years we have seen a number of major global new media firms attract negative headlines when significant numbers of their personal detail harvested, in some cases without their knowledge. Facebook found its platform was used by a Cambridge University academic who distributed a 'personality test' app on the social media platform. Extensive personal data/details of over 87 million users was collected by the app. The academic who distributed the original 'personality test' application then passed the personal data collected to a firm called Cambridge Analytica. The data was then used in online election advertising in the 2016 US presidential election.

The Facebook and Google+ data breaches represent significant and high profile personal data breaches by major new media players. EU regulators are still investigating these cases and we await the findings from the relevant enquiries. What these and other similar cases demonstrate is how data breaches can occur and the data harvested potentially used to influence general elections. It is likely that some data subjects have become more sensitive to potential personal data breaches when choosing the new media services they interact with in future. As such, privacy seals have potential significant role to play in differentiating between competing application/service providers. For example, as an EU member state data subject you have a choice of three similar online service providers to choose from. All three-service providers must under EU data protection rules adhere to the GDPR. However, only one of the three online service providers displays an independently verified privacy seal. Privacy sensitive data subjects may choose to select the service provider who displays the seal. It is important that the new media industry adopts initiatives to reassure consumers that their data is safe. As Akerlof (1978) warns, the inability of consumers to differentiate between scrupulous and unscrupulous sellers risk undermining the whole industry and this can lead to market failure.

5.3.2 Implications for Media Regulators

Data protection regulators may well view these results positively i.e. the presence of the privacy seal does not negatively impact on data processors. Furthermore, it is possible that over time, as the use of GDPR compliance seals become more widely used, data subjects may only engage with new media firms (i.e. data controllers) who display privacy seals (and/or similar 'privacy' kite marks). We may see a situation where data subjects seek out firms who display privacy seals much in the same way as people seek out efficient energy labels when purchasing domestic white goods. We do already see some evidence to support this. For example, Microsoft ran a global campaign promoting their 'cloud' based services as already being GDPR compliant in advance of the GDPR effective date. Apple have done something similar, as they are primarily a hardware seller, they are differentiating themselves from those media firms who are reliant on making data subjects personal details available to maximise advertising revenues.

Another conclusion of these experiments is that the so-called 'privacy paradox' is still alive and well. The findings here again show while respondents state that specific categories of data should be treated as 'sensitive' personal data yet they are willing to disclose this information in return for a relatively small amount of money. Perhaps as data subjects become more aware of how their personal data is collected and used by new media firms, they will adjust their behaviour and we may see the privacy paradox type behaviour reduced or disappear.

From a privacy seal perspective, media regulators should look to the success of energy labels and adopt a similar approach. Energy labels took a complex technical area and produced a standardized and visually easy to interpret approach to its labels. As well as helping consumers to understand complex technical information, the approach also encouraged manufacturers to produce more energy efficient goods. We could see privacy seals enjoy a similar success. The presence of a standardized seal would allow media consumers to easily see how media firms manage consumers personal data. It would also allow consumers to choose between scrupulous media firms who have had their data practices independently verified, from those who perhaps do not comply with

data protection rules. As Akerlof (1978) warns, without this ability to differentiate, market may fail. A markets failure has the potential to be bad for consumers and media firms as well as wider negative impact on economic growth.

Regulators need to expand the list of sensitive data categories to reflect the modern era. The survey findings here in relation to RO2 clearly demonstrate that the actual list of sensitive data categories in the GDPR does not meet with consumers expectations of what should be classed as sensitive.

5.4 Original Contribution to Knowledge

This thesis makes a number of original contributions to knowledge. Phillips and Pugh (2010) identify 15 different definitions of originality and this thesis makes contributions in the following ways:

Carrying out empirical work that hasn't been done before:

The surveys carried out in relation to RO2 represent original empirical research that specifically addresses respondents' views on what they believe *should* be considered 'sensitive' as data categories.

Additionally, there is no existing research (empirical or otherwise) that compares EU and US respondents' views in relation to sensitive personal data categories.

Making a synthesis that has not been made before:

Existing research in relation to privacy seals has been conducted exclusively on US based respondents. In this thesis a survey was used to identify what categories of personal data EU and US respondents considered most sensitive. These results were then used to conduct experiments on both EU member and US state residents to examine what impact the presence of privacy seals have on personal information disclosure. The results also allowed for a comparison of the results from the respondents based in different jurisdictions.

New evidence on an old issue:

Existing research on privacy seals and related privacy iconography has relied on icons created by the authors. This brings obvious methodological issues in relation to their design etc. In the research conducted here actual 'real world' operation privacy seals run by private companies have been used in the experiments and this is an original approach. This extends our knowledge on how contextual clues impact on personal information disclosure. The findings also make a contribution to confirm existing theoretical literature on the presence of the so called 'privacy paradox'. In the surveys conducted for this thesis, participants confirmed the categories of data that they thought should

have added 'sensitive data' protection in law. However, in the experiments the vast majority disclosed many of these personal data categories for less than \$1.

5.5 Concluding Remarks

Technological developments have facilitated the rapid growth of a global new media industry. For some firms like Facebook, Apple and Google, first mover advantage has resulted in them capturing very significant percentages of global market share. Technology now allows sophisticated advertising to media consumers based on their personal data disclosures has seen a small number of new medial firms dominate this sector, with some commentators referring to it as an example of monopoly capitalism.

In May 2017, the editorial in the Economist declared the world most valuable recourse was no longer oil, but data. It went on to call for this new media sector to have more regulation. Apple's chief executive Tim Cook echoed this by declared privacy to be one of the most important issues of the 21st century and called for greater regulation of the industry to protect the wider interests of society (Thornhill 2018). Cook has even described the collection of personal data as surveillance (Lomas 2018).

Perhaps we should not be overly surprized with the warning from Tim Cook. In the 1970's we heard Irwin Altman warn that individual privacy was required for cultural survival.

Additionally, in light of the Cambridge Analytica scandal in March 2018, Facebook has announced measures to increase the transparency around who places political advertisements on Facebook (Hern and Waterson, 2018). It remains to be seen now effective this increased transparency is going to be.

The 2018 GDPR represents an attempt by EU regulators to address the concerns about how technology can gown in a way that helps wider society and thus avoid the somewhat dystopian cultural and democratic harms that Cook and Altman warn us about. Although its existence emanates from an EU

regulation, the 'Privacy Shield' arrangement for international data transfers means that the GDPR has effectively become the global standard for data protection.

The effective enforcement of meaningful penalties for breaches of regulatory rules, has an important role to play in protecting data subjects personal data. However, in the case of large new media firms, regulation and deterrents might not be enough. Some commentators suggest that for a firm like Facebook,

“the nature and scale of its operations make it nearly impossible to avoid major data breaches that expose highly personal data” (Coldewey 2018).

If we accept that new media firms (as data controllers) cannot guarantee to keep our personal data safe, then it is important that consumers of new media (i.e. data subjects) take responsibility for who they decide to share their personal data with. While privacy seals are not the only weapons that can be deployed to protect against the 'data industrial complex', an effective privacy seal accreditation scheme can be an important tool for both data subjects and data controllers to engender trust in each other.

Mai et al (2010) conducted research into how the use of privacy seals by online vendors impacting on consumers purchasing behaviour. The authors usefully differentiate between what they call 'scrupulous' and 'unscrupulous' vendors. Scrupulous vendors are those implement and monitor stringent privacy policies and incur additional costs because of this. Unscrupulous vendors do not incur these costs. The research adopts the premise that when making purchases online, consumers face a choice between buying from privacy seal bearing vendors or from non-privacy seal bearing vendors. The latter is associated with an enhanced risk of privacy violation for the purchaser (data subject). There is almost no way for potential consumers to distinguish between the 'scrupulous' and 'unscrupulous' vendors and therefore privacy seals potentially have a role here.

It is my belief that it is data subjects themselves who need to take greater responsibility for what personal data we choose to share with specific new

media service providers. Third party verified privacy seals can have a role in helping data subject select new media services who adhere to the relevant regulatory requirements. In the UK we have an OFSTEAD system for independently evaluating schools performance against a set of national criteria. This does not make all schools excellent but it does help with decision making for parents. A similar independent third-party evaluation of the data collection and sharing of media and large tech firms is needed to help us as consumers (i.e. data subjects) decide who we interact with.

In the 1980's, HIV was considered an issue of national health and in the UK a national public information campaign to inform the public about the potential risks and how to take steps to avoid those risks. For me, the data industrial complex and new media and tech firms ability to weaponize our data against us for either commercial or political reasons, represents a serious threat that requires a national/international campaign to inform and mitigate against the most serious consequences. Third party accredited privacy seals have a role to play in protecting data subjects. It is not the sole solution, but it has a role to play.

Some commentators have warned that data protection regulators have been, "corrupted and compromised through timidity and neglect" (Davies 2001, pg289). Davies does acknowledge that some may interpret his views as 'extreme' and/or alarmist. However, he defends his position pointing out that once a fundamental right has been established, the rigorous protection of these rights is in fact a conservative notion and this it is the transgressor who becomes radical. He believes both the government and the private sector conveniently invert this notion and it is those who call for greater privacy as radical rather than a conservative notion.

As Lucy Kung notes, the new media industry has faced many new innovations over the past 100 years. There is one persistent pattern that all CEO's should be aware of, "technology gives, and technology takes away" (Küng 2008, p.125).

"Under such circumstances, dialogue between media management scholars and policy makers, as mentioned, is critical" (Rohn 2018, p.436).

Media management scholars can be used to identify emerging trends in the sector that may give firms a strategic advantage over competitors, rather than just waiting on regulators to publish compliance rules.

Google (Alphabet), Facebook and Microsoft have all established dominant positions in the new media sector. This has mainly been driven through their ability to innovate and capitalize on technological advances (Doyle 2013). However, their ability to retain these dominant positions may depend on their ability to react to their users desire not just for innovations in terms of functionality and features, but also in their ability to react to concerns. As new media consumers become more aware of how their personal data is shared/traded etc, there may well be a role for third-party verified/accredited Privacy Seals in giving consumers the reassurances they need that data controllers are trustworthy. As the new media sector further evolves, retaining dominant positions within the sector may depend on such assurances/innovations.

5.6 Limitations:

The methodological limitations associated with surveys and experimental methods as well as recruitment of participants using MTurk are set out with in Chapter 3. A key limitation in relation to RO3 experimental findings is that participants did not see or understand what the privacy seals displayed in some of the treatments. There are methodological reasons as to why experiment participants are not explicitly alerted to the presence of the privacy seal.

5.7 Future Research:

The experiment and surveys in this thesis were conducted in 2016 and 2017. In 2018 we have seen controversies in terms of how personal information was harvested from Facebook and used in political advertising during the US presidential and Brexit referendum in the UK. It would be interesting to see if

these controversies have made people more privacy aware in terms of personal information disclosure.

“The tremendous speed of change in the media industries and practices calls for conducting more longitudinal and time series studies to avoid only seeing screenshots of reality that do not capture all the trends and full complexity” (Rohn 2018, p.437).

It will be interesting to see the results of the experiments conducted for this research change over time as public becomes more aware of how their personal information is used in behavioural advertising and online political campaigns etc.

The findings here could be used as a longitudinal study to establish how attitudes to privacy, and specifically privacy seal's impact on personal information disclosure, develops over time.

References

Acquisti, A., 2004, May. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce* (pp. 21-29).

Acquisti, A. and Grossklags, J., 2005. Privacy and rationality in individual decision making. *IEEE security & privacy*, 3 (1), 26-33.

Acquisti, A. and Grossklags, J., 2007. What can behavioral economics teach us about privacy. *Digital privacy: theory, technologies and practices*, 18, 363-377.

Acquisti, A. and Grossklags, J., 2012. An online survey experiment on ambiguity and privacy. *Communications & Strategies*, (88), 19-39.

Acquisti, A., John, L.K. and Loewenstein, G., 2013. What is privacy worth?. *The Journal of Legal Studies*, 42(2), 249-274.

Acquisti, A., Taylor, C. and Wagman, L., 2016. The economics of privacy. *Journal of Economic Literature*, 54(2), 442-92.

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K. and Wetzels, M., 2015. Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of retailing*, 91(1), 34-49

Akerlof, G. A., 1978. The market for "Lemons": quality uncertainty and the market mechanism. In: Diamond, P. and Rothschild, M., eds. *Uncertainty in Economics*. New York: Academic Press, 235-251.

Albarran, A.B., 2004. Media economics. *In*: Dowling, J.D.H, McQuail, D., Schlesinger, P., Wartella, E., eds. *The SAGE handbook of media studies*. London, Sage Publications, 291-308.

Alphabet Investor Relations, 2018. *Alphabet Announces Fourth Quarter and Fiscal Year 2017 Results* [online]. Available from: https://abc.xyz/investor/static/pdf/2017Q4_alphabet_earnings_release.pdf?cache=33ec3b1 [Accessed 25 October 2018]

Altman, I., 1975. *The Environment and Social Behavior: Privacy – Personal Space – Territory – Crowding*. California: Wadsworth Publishing Company.

Altman, I., 1977. Privacy Regulation: Culturally Universal or Culturally Specific? *Journal of Social Issues*, 33 (3), 66-84 .

Angner, E., 2015. To navigate safely in the vast sea of empirical facts. *Synthese*, 192 (11), 3557-3575.

Arsenault, A.H., 2017. The datafication of media: Big data and the media industries. *International Journal of Media & Cultural Politics*, 13(1-2), 7-24.

Arthur, C., 2012. The end of online privacy?. *The Guardian* [online] 28th February 2012. Available from: <http://www.guardian.co.uk/technology/2012/feb/28/the-end-of-online-privacy?INTCMP=SRCH> [Accessed 01 August 2012]

Arthur, C., 2013. Tim Berners-Lee: UK and US must do more to protect users privacy *BBC News* [online], 22 November 2013. Available from:

<https://www.theguardian.com/technology/2013/nov/22/tim-berners-lee-internet-privacy-surveillance-censorship> [Accessed 14 Feb 2014]

Article 29 Working Party, 2011. *Advice paper on special categories of data “sensitive data”* [online]. Brussels: European Commission, Brussels Ref. Available from: [http:// https://ec.europa.eu/newsroom/article29/news-overview.cfm](http://https://ec.europa.eu/newsroom/article29/news-overview.cfm) [Accessed 22 December 2015]

Aust, F., Diedenhofen, B., Ullrich, S. and Musch, J., 2013. Seriousness checks are useful to improve data validity in online research. *Behavior research methods*, 45 (2), 527-535.

Awad, N. F., and Krishnan, M. S., 2006. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiles Online for Personalization. *MIS Quarterly*, 30 (1), 13-28.

Athique, A., 2018. The dynamics and potentials of big data for audience research. *Media, Culture & Society*, 40 (1), 59-74.

Barnett, E., 2010. Facebook’s Mark Zuckerberg says privacy is no longer a ‘social norm’. *Telegraph* [online], 11 Jan 2010. Available from: <http://www.telegraph.co.uk/technology/facebook/6966628/Facebooks-Mark-Zuckerberg-says-privacy-is-no-longer-a-social-norm.html> [Accessed 24 March 2013]

Beck, J. C., 2001. Get a grip! Regulating cyberspace won’t be easy. *Business Law Today*, (Issue 14), 14-19.

Bertoldi, P. and Atanasiu, B., 2007. Electricity consumption and efficiency trends in the enlarged European Union – Status Report 2006. *IES–JRC. European Union*. Available from: <https://core.ac.uk/reader/38617836> [Accessed 16 July 2016]

Biggam, R., 2015. The challenges for public policy of adjusting to a multi-platform environment. *Journal of Media Business Studies*, 12 (1), 89-102.

Brandimarte, L, Acquisti, A., and Loewenstein, G., 2013. Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science*, 4 (3), 340-347.

Bryman, A., 2012. *Social Research Methods*. 4th edition. Oxford: Oxford University Press.

Bryman, A. and Bell, E., 2003. *Business Research Methods*. Oxford: Oxford University Press.

Buhrmester, M., Kwang, T. and Gosling, S.D., 2011. Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on psychological Science*, 6 (1), 3-5.

Cadwalladr, C., 2016. Google, democracy and the truth about internet search. *The Observer* [online]. 4 December 2016. Available from: https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook?CMP=share_btn_tw [Accessed 4 October 2017]

Carolan, E. and Castillo-Mayen, M.R., 2014. Why More User Control Does Not Mean More User Privacy: An Empirical (and Counter-Intuitive) Assessment of European E-Privacy Laws. *Va. JL & Tech.*, 19, 324.

Cate, F. H., 1999. The Changing Face Of Privacy Protection In The European Union And The United States. *Indiana Law Review*, 33,173.

Cate, F. H., 2011. A Transatlantic Convergence on Privacy. *IEEE Computer and Reliability Societies*, 9 (1), 76-79.

Cicourel, A. V., 1964. *Method and Measurement in Sociology*. New York: Free Press.

Clarke, R., 2013. Information Privacy. *RogerClarke.com* [online]. Available from: <http://www.rogerclarke.com/DV/Intro.html#InfoPriv> [Accessed 20 April 2014]

Clifford, S., Jewell, R.M. and Waggoner, P.D., 2015. Are samples drawn from Mechanical Turk valid for research on political ideology?. *Research & Politics*, 2 (4), 1-9.

Coldewey, D., 2018. Facebook can't keep you safe. *TechCrunch* [online], 10 June 2018. Available from: <https://techcrunch.com/2018/10/01/facebook-cant-keep-you-safe/> [Accessed 15 Oct 2018]

Constine, J., 2018. Worries linger as Facebook withholds stolen searches & checkins. *Techcrunch* [online]. 15 October 2018. Available from: https://techcrunch.com/2018/10/15/facebook-breach-searches-locations/?utm_source=tctwreshare&sr_share=twitter [Accessed 17 October 2018]

Cook, T.D. and Campbell, D.T., 1979. The design and conduct of true experiments and quasi-experiments in field settings. *In* Mowday, R.T., Steers, R.M., eds, Reproduced in part in *Research in Organizations: Issues and Controversies*. California: Goodyear Publishing Company.

Couldry, N. and Turow, J., 2014. Advertising, Big Data, and the clearance of the public realm. *International Journal of Communication*, 8 (1), 1710-1726.

Council of Europe, 1950. The European Convention on Human Rights. *Council of Europe* [online] 4 November 1950. Available from: <http://www.hri.org/docs/ECHR50.html> - C.Art8 [Accessed 24 March 2010]

Court of Justice of the European Union, 2015. Judgment in Case C 362/14 Maximilian Schrems v Data Protection Commissioner. Press Release No 117/15, Luxembourg, 6 Oct 2015. Available from: <http://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117en.pdf> [Accessed 16 March 2016]

Davies, S., 2001. Unprincipled Privacy: Why the foundations of Data Protection are failing us. *University of South Wales Law Journal*, 24 (1). 284-289.

Diez, D. M., Barr, D. D., and Cetinkaya-Dundel, M., 2012. *OpenIntro Statistics*, 2nd edition. Creative Commons. [Openintro.org](http://openintro.org).

Dillman, D.A., 2000. Procedures for conducting government-sponsored establishment surveys: Comparisons of the total design method (TDM), a traditional cost-compensation model, and tailored design. *In Proceedings of*

American Statistical Association, Second International Conference on Establishment Surveys, 343-352.

Dillon, J. and Wals, A.E., 2006. On the danger of blurring methods, methodologies and ideologies in environmental education research. *Environmental Education Research*, 12(3-4), 549-558.

Directive 95/46/EC, 1995. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

Brussels, European Commission. Available from:

<http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML> [Accessed 11 August 2012]

DoC, 2010. *Commercial Data Privacy and Innovation in the Internet Economy*. US Department of Commerce [online]. Washington DC: US Department of Commerce. Available from:

https://www.ntia.doc.gov/files/ntia/publications/iptf_privacy_greenpaper_12162010.pdf [Accessed 15 August 2012]

Doyle, G., 2013. *Understanding media economics*. 2nd edition. London: Sage.

Doyle, G., 2018. Television and the development of the data economy: Data analysis, power and the public interest. *International Journal of Digital Television*, 9(1), 53-68.

Downs, J.S., Holbrook, M.B., Sheng, S. and Cranor, L.F., 2010, April. Are your participants gaming the system?: screening mechanical turk workers. In

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2399-2402.

Druckman, J.N. and Nelson, K.R., 2003. Framing and deliberation: How citizens' conversations limit elite influence. *American Journal of Political Science*, 47 (4), 729-745.

Dupagne, M., 2018. Methodological Approaches in Media Management and Economics. In Albarran, A., Meirzejewska, B., Jung, J., eds, *Handbook of Media Management and Economics*. New Jersey: Routledge, 363-378.

Economist, 2017. The world's most valuable resource is no longer oil, but data. *Economist*[online], 6 May 2017. Available from: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data> [Accessed 02 November 2018]

Edwards, L. and Abel, W., 2014. The use of privacy icons and standard contract terms for generating consumer trust and confidence in digital services. *CREATE working paper series*. Glasgow: CREATE.

ePrivacy, 2018. E-privacyseal. *ePrivacy*[online]. Hamburg: ePrivacy. Available from: <https://www.eprivacy.eu/en/privacy-seals/eprivacyseal/> [Accessed 11 Nov 2018]

European Commission, 2010. *A comprehensive approach on personal data protection in the European Union*[online]. Brussels, European Commission. Available from: https://edps.europa.eu/data-protection/our-work/publications/opinions/comprehensive-approach-personal-data-protection_en [Accessed 15 August 2012]

European Commission, 2012. *Commission proposes a comprehensive reform of the data protection rules* [online]. Brussels, European Commission. Available from: http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm [Accessed on: 18 August 2012]

EuroPriSe, 2017. Website Privacy Certification. *EuroPriSe* [online]. Available from: <https://www.european-privacy-seal.eu/EPS-en/website-privacy-certification-overview> [Accessed on: 11 November 2018]

Evens, T. and Van Damme, K., 2016. Consumers' willingness to share personal data: Implications for newspapers' business models. *International Journal on Media Management*, 18 (1), 25-41.

Facebook, 2019. Facebook Company Info: Stats. *Facebook Newsroom* [online]. Available from: <https://newsroom.fb.com/company-info/> [Accessed 04 April 2018]

Farahbakhsh, R., Mohammadi, S., Han, X., Cuevas, A. and Crespi, N., 2017, August. Evolution of publicly disclosed information in Facebook profiles. In *2017 IEEE Trustcom/BigDataSE/ICSS*, 9-16.

Falmer, G., 2015. What you need to know about ICO Privacy Seals. *Information Commissioner's Office Blog* [online]. 28 Jan 2015. Available from: <https://iconewsblog.wordpress.com/2015/01/28/what-you-need-to-know-about-ico-privacy-seals/> [Accessed 12 October 2016]

Feri, F., Giannetti, C. and Jentzsch, N., 2016. Disclosure of personal information under risk of privacy shocks. *Journal of Economic Behavior & Organization*, 123, 38-148.

Fisher, R.A., 1922. On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, Issue 85 (1), 87-94.

Fried, C., 1968. Privacy. *The Yale Law Journal*, Vol. 77, 475-493.

FTC, 2010. Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers. *FTC* [online]. December 2010. Available from: <http://www.ftc.gov/reports/preliminary-ftc-staff-report-protecting-consumer-privacy-era-rapid-change-proposed-framework> [Accessed 15 August 2012]

FTC, 2012. Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers. *FTC* [online]. March 2012. Available from: <http://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers> [Accessed on: 05 February 2014]

Gaines, B.J., Kuklinski, J.H. and Quirk, P.J., 2007. The logic of the survey experiment reexamined. *Political Analysis*, 15(1), 1-20.

GDPR, 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council. *Official Journal of the European Union*, 27 April 2016. Available from: <https://gdpr-info.eu/art-4-gdpr/> [Accessed 04 June 2016]

Gelles, D., Bradshaw, T., and Palmer, M., 2009. Change the rules and watch users log off. *Financial Times*, 12/13 Dec 2009.

Goldfarb, A. and Tucker, C.E., 2011. Privacy regulation and online advertising. *Management science*, 57 (1), 57-71.

Granados, N. F., Gupta, A., and Kauffman, R. J., 2006. The Impact of IT on Market Information and Transparency: A Unified Theoretical Framework. *Journal of the Association for Information Systems*, Issue 7 (3).

Greenwald, G., 2013. Revealed: how US and UK spy agencies defect internet privacy and security. *The Guardian* [online], 06 September 2013. Available from: <https://www.theguardian.com/world/2013/sep/05/nsa-gchq-encryption-codes-security> [Accessed 22 December 2013]

Grunert, K.G. and Wills, J.M., 2007. A review of European research on consumer response to nutrition information on food labels. *Journal of Public Health*, 15 (5), 385-399.

Ha, A., 2016. Edward Snowden says 'the central problem of the future' is control of user data'. *Techcrunch* [online]. 13 December 2016. Available from: <https://techcrunch.com/2016/12/13/edward-snowden-says-the-central-problem-of-the-future-is-control-of-user-data/> [Accessed 7 January 2017]

Harcourt, A., 2005. *The European Union and the Regulation of Media Markets*. Manchester: Manchester University Press.

Henderson, S. C. and Snyder, C.A., 1999. Personal information privacy: implications for MIS managers. *Information & Management*, (Issue 36), 213-220.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. and McElreath, R., 2001. In search of homo economicus: behavioral experiments in 15 small-scale societies. *The American Economic Review*, 91(2), 73-78.

Hern, A., and Pegg, D., 2018. Facebook fined for data breaches in Cambridge Analytica scandal. *The Guardian* [online], 11 July 2018. Available from: <https://www.theguardian.com/technology/2018/jul/11/facebook-fined-for-data-breaches-in-cambridge-analytica-scandal> [Accessed 11 July 2018]

Hern, A., and Waterson, J., 2018. Facebook cracks down on 'dark ads' by British political groups. *The Guardian* [online], 16 October 2016. Available from: <https://www.theguardian.com/technology/2018/oct/16/facebook-dark-ads-british-political-groups> [Accessed 18 October 2018]

Hixson, R.F., 1987. *Privacy in a public society: Human rights in conflict*. New York: Oxford University Press.

Hochheiser, H., 2002. The platform for privacy preference as a social protocol: An examination within the US policy context. *ACM Transactions on Internet Technology (TOIT)*, 2 (4), 276-306.

Hoofnagle, C.J., Soltani, A., Good, N. and Wambach, D.J., 2012. Behavioral advertising: The offer you can't refuse. *Harv. L. & Pol'y Rev.*, 6, 273.

Horiuchi, Y., Imai, K. and Taniguchi, N., 2007. Designing and analyzing randomized experiments: Application to a Japanese election survey experiment. *American Journal of Political Science*, 51(3), pp.669-687.

Horst, S.O. and Murschetz, P.C., 2019. Strategic media entrepreneurship: Theory development and problematization. *Journal of Media Management and Entrepreneurship (JMME)*, 1(1), 1-26

Hovland, C.I. and Weiss, W., 1951. The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15 (4), 635-650.

ICO, 2012. *Guidance on the rules and use of cookies and similar technologies*. UK Information Commissioner Office[online]. May 2012. Available from: https://ico.org.uk/media/1545/cookies_guidance.pdf [Accessed 29 September 2012]

ICO, 2016. *Improve your practices*. Information Commissioner Office blog [online]. Available from: <https://ico.org.uk/for-organisations/improve-your-practices/privacy-seals/> [Accessed on: 14 September 2016]

Jayakar, K., 2018. Media Policy. In *Handbook of Media Management and Economics*. In Albarran, A., Meirzejewska, B., Jung, J., eds, *Handbook of Media Management and Economics*. 2nd edition. New Jersey: Routledge, New York: Routledge.

Jensen, C. and Potts, C., 2004. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 471-478.

Jensen, C., Potts, C. and Jensen, C., 2005. Privacy practices of Internet users: self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63 (1), 203-227.

John, L.K., Acquisti, A. and Loewenstein, G., 2010. Strangers on a plane: Context-dependent willingness to divulge sensitive information. *Journal of consumer research*, 37 (5), 858-873.

Jung, J., Mierzejewska, B. and Albarran, A., 2018. *Handbook of Media Management and Economics*. 2nd edition New York: Routledge.

Kahneman, D., 2003. Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, 93 (5), 1449-1475.

Kidder, L. and Judd, C., 1986. *Research methods in social science*. New York: CBS College.

Kirwan, P., 2012. EU cookie law: stop whining and just get on with it. *Wired* [online], 24 May 2012. Available from: <http://www.wired.co.uk/news/archive/2012-05/24/eu-cookie-law-moaning> [Accessed 05 September 2012]

Klein, J.T. and Newell, W.H., 1997. Advancing interdisciplinary studies. *Handbook of the undergraduate curriculum: A comprehensive guide to purposes, structures, practices, and change*, pp.393-415.

Koetsier, J., 2018. Digital Duopoly Declining? Facebook's, Google's Share Of Digital Ad Dollars Dropping. *Forbes* [online]. 19 March 2018. Available from: <https://www.forbes.com/sites/johnkoetsier/2018/03/19/digital-duopoly-declining-facebooks-googles-share-of-digital-ad-dollars-dropping/> [Accessed 12 November 2018]

Korolova, A., 2010. Privacy violations using microtargeted ads: A case study. In *2010 IEEE International Conference on Data Mining Workshops*. 474-482.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., 2002. The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice? *Public Opinion Quarterly*, (issue 66), 371-403.

Küng, L., 2017. Reflections on the ascendancy of technology in the media and its implications for organisations and their leaders. *The Journal of Media Innovations*, 4 (1), 77-81.

Küng, L., 2018. *Strategic management in the media: Theory to practice*. London: Sage.

Küng, L., Picard, R.G. and Towse, R. eds., 2008. *The internet and the mass media*. London: Sage.

Kunreuther, H., Ginsberg, R., Miller, L., Sagi, P., Slovic, P., Borkan, B., and Katz, N., 1978. *Disaster Insurance Protection: Public Policy Lessons*. New York: Wiley.

Lang, C., 2018. 'It's Not Good.' Mark Zuckerberg Discusses the #DeleteFacebook Campaign. *Time* [online], 28 March 2018. Available from:

<http://time.com/5210799/mark-zuckerberg-addresses-delete-facebook-campaign-after-cambridge-analytica/> [Accessed 14 March 2019]

Leadbeater, C., 2010. Cloud computing: how information giants are setting the pace for the internet's next decade. *The Guardian* [online], 7 Feb 2010. Available from: <http://www.guardian.co.uk/technology/2010/feb/07/cloud-computing-google-apple> [Accessed 23 March 2013]

Lee, D., 2013. EE defends user-data selling scheme following police interest. *BBC News* [online], 13 May 2013 Available from: <http://www.bbc.co.uk/news/technology-22510792> [Accessed 15 May 2013]

Lloyd, I. J., 2011. *Information technology law*. 6th edition. London: Oxford University Press.

Lomas, N., 2018. Apple's Tim Cook makes blistering attack on the 'data industrial complex'. *Techcrunch* [online], 24 October 2018. Available from: <https://techcrunch.com/2018/10/24/apples-tim-cook-makes-blistering-attack-on-the-data-industrial-complex/> [Accessed 24 October 2018]

Loring, T. B., 2002. Analysis of the Information Privacy Protection Afforded by the European Union and the United States. *Texas International Law Journal*, 37, 421-460.

Madden, M. and Rainie, L., 2015. *Americans' attitudes about privacy, security and surveillance*. Washington: Pew Research Center.

Mai, B., Menon, N.M. and Sarkar, S., 2010. No free lunch: Price premium for privacy seal-bearing vendors. *Journal of Management Information Systems*, 27 (2), 189-212.

Margulis, S.T., 2003. On the status and contribution of Westin's and Altman's theories of privacy. *Journal of Social Issues*, 59 (2), 411-429.

Maxeiner, J. R., 1995. Freedom of Information and the EU Data Protection Directive. 48 FED. COM. L.J. 93, 95. Available from: <http://www.law.indiana.edu/fclj/pubs/v48/no1/maxeiner.html> [Accessed 11 August 2014]

McDonald, A.M. and Cranor, L.F., 2008. Cost of reading privacy policies. *ISJLP*, 4, 543.

McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Baltimore: Sparky House Publishing.

Mclaughlan, M., 2017. Personal data for sale on 'huge scale' warns consumer group. *The Scotsman* [Online]. Available from: <http://www.scotsman.com/news/uk/personal-data-for-sale-on-huge-scale-warns-consumer-group-1-4344881> [Accessed 25 January 2017]

Metz, C., and Goodin, D, 2010. Google: Street View spycars did slurp your Wi-Fi. *The Register* [online]. 14 May 2010. Available from: http://www.theregister.co.uk/2010/05/14/google_street_view_cars_were_collecting_payload_data_from_wifi_networks/ [Accessed 18 May 2010]

Mierzejewska, B. and Shaver, D., 2014. Key changes impacting media management research. *International Journal on Media Management*, 16, 47-54.

Milne, G.R. and Culnan, M.J., 2002. Using the content of online privacy notices to inform public policy: A longitudinal analysis of the 1998-2001 US Web surveys. *The Information Society*, 18(5), pp.345-359.

Miyazaki, A.D. and Krishnamurthy, S., 2002. Internet seals of approval: Effects on online privacy policies and consumer perceptions. *Journal of Consumer Affairs*, 36(1), pp.28-49.

Murgia, M., 2017. Facebook fined €110 by European Commission over WhatsApp deal. *Financial Times* [online], 18th May 2017. Available from: <https://www.ft.com/content/a2dad48-3bb1-11e7-821a-6027b8a20f23?mhq5j=e1> [Accessed on: 20 July 2017]

Murray, P. J., 1998. The Adequacy Standard Under Directive: 95/46/EC: Does U.S. Data Protection Meet This Standard? *Fordham International Law Journal*, Volume 21, Issue 3, pp 932-1018.

Napoli, P.M. and Roepnack, A., 2018. Big Data and Media Management. In Albarran, A., Mierzejewska, B., Jung, J., eds, *Handbook of Media Management and Economics*. 2nd edition. New York: Routledge, 410-421.

Noam, E.M., 1997. Privacy and self-regulation: Markets for electronic privacy. *Privacy and Self-Regulation in the Information Age*, 21-33.

Norberg, P.A., Horne, D.R. and Horne, D.A., 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer Affairs*, 41 (1), 100-126.

O'Kane, C., 2015. A View from the Amsterdam Privacy Conference 2015. *CREATE* [online], 23 October 2015. Available from: <http://www.create.ac.uk/blog/2015/11/05/a-view-from-the-amsterdam-privacy-conference-2015/> [Accessed 16 March 2016]

O'Kane, C., 2018. As Facebook admits grabbing personal data of 87m users, it's time we switched off these i-spies. *Mirror* [online], 6 April 2018. Available from: <https://www.mirror.co.uk/news/uk-news/facebook-admits-grabbing-personal-data-12312131> [Accessed 16 Feb 2018]

Oliver, J., 2014. Dynamic capabilities and superior firm performance in the UK media industry. *Journal of Media Business Studies*, 11 (2), 57-78.

Oliver, J.J., 2018. Re-evaluating the role of scenario planning. *Business Horizons*, 61 (2), 339-352.

Oliver, J.J., 2018. Strategic transformations in the media. *Journal of Media Business Studies*, 15 (4), 278-299.

Oliver, J.J. and Picard, R.G., 2020. Shaping the corporate perimeter in a changing media industry. *International Journal on Media Management*, 1-16.

Oppenheimer, D.M., Meyvis, T. and Davidenko, N., 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45 (4), 867-872.

Oxford Dictionaries, 2018. *Oxford Dictionaries* [online]. Available from: https://en.oxforddictionaries.com/definition/new_media [Accessed 11 September 2018]

Palen, L. and Dourish, P., 2003. Unpacking privacy for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 129-136.

Paolacci, G., Chandler, J. and Ipeirotis, P.G., 2010. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5 (5), 411-419.

Peer, E., Vosgerau, J. and Acquisti, A., 2014. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk'. *Behavior Research Methods*, 46 (4), 1023-1031.

Peltzman, S., 1975. The effects of automobile safety regulation. *Journal of political Economy*, 83 (4), 677-725.

Phillips, D. L., 1973. *Abandoning Method*. San Francisco: Jossey-Bass.

Phillips, E. and Pugh, D., 2010. *How to get a PhD: A handbook for students and their supervisors*. London: McGraw-Hill Education.

- Picard, R. G., 1989. *Media Economics: Concepts and Issues*. California: Sage.
- Picard, R. G., 2002. *The Economics and Financing of Media Companies*. New York: Fordham University Press.
- Posner, R. A., 1981. The Economics Of Privacy. *The American Economic Review*, 71 (2), 405-409.
- Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S. and Dabbish, L., 2013. Anonymity, privacy, and security online. *Pew Research Center*, 5.
- Rand, D.G., 2012. The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of theoretical biology*, 299, 172-179.
- Reed, C., 2011. *Computer Law*. Seventh Edition. Oxford: Oxford University Press.
- Regulation, P., 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council. REGULATION (EU), 679.
- Reidenberg, J.R., Breaux, T., Cranor, L.F., French, B., Grannis, A., Graves, J.T., Liu, F., McDonald, A., Norton, T.B. and Ramanath, R., 2015. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Tech, LJ*, (Issue 30), 39.

Rifon, N.J., LaRose, R. and Choi, S.M., 2005. Your privacy is sealed: Effects of web privacy seals on trust and personal disclosures. *Journal of Consumer Affairs*, 39 (2), 339-362.

Roscoe, J.T., 1975. *Fundamental research statistics for the behavioural sciences*. New York: Holt, Rineheart and Winston.

Rohn, U., 2018. Media Management Research in the Twenty-First Century. In Albarran, A., Meirzejewska, B., Jung, J., eds, *Handbook of Media Management and Economics*. 2nd Edition. New York: Routledge, 525-441.

Rosenthal, R., and Jacobson, L., 1968. Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development. New York: Holt, Rineheart & Winston.

Rossi, B., 2017. GDPR compliance – the real implications for businesses. *Information Age* [online], 14 June 2017. Available from: <http://www.information-age.com/gdpr-compliance-real-implications-businesses-123466772/> [Accessed 25 July 2017]

Scheuren, F., 2004. What is a survey? *American Statistical Association* [online]. Available from: http://www.parkdatabase.org/files/documents/2004_What-is-a-Survey_ASA_F-Scheuren.pdf [Accessed 17 October 2017]

Schutz, A., 1962. Concept and theory formation in the social sciences. In *Collected Papers I*. Dordrecht: Springer, 48-66.

Sekaran, U., 1992. *Research Methods for Business – A skill building approach*. 2nd edition. New York: John Wiley & Sons.

Selten, R., 1999. What is Bounded Rationality?. SFB Discussion Paper B-454. Dahlem Conference, May 1999. Available from: <http://www.wiwi.uni-bonn.de/sfb303/papers/1999/b/bonnsfb454.pdf> [Accessed 01 May 2014]

Shaffer, G., 2000. Globalization and Social Protection: The Impact of EU and International Rules in the Ratcheting Up of U.S. Privacy Standards. *Yale Journal of International Law*, (Issue 25), 1.

Shafir, E. and Mullainathan, S., 2013. *Scarcity: Why having too little means so much*. New York: Times Books.

Simon, H.A., 1986. Rationality in psychology and economics. *Journal of Business*, 209-224.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J., 2006. Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, (Issue 70), 66-77.

Solove, D.J., 2007. I've got nothing to hide and other misunderstandings of privacy. *San Diego Law Review*, (Issue 44), 745.

StatCounter, 2019. Search Engine Market Share Worldwide: March 2019-March 2019. *Statcounter* [online] Available from: <http://gs.statcounter.com/search-engine-market-share> [Accessed 10 April 2019]

Statista, 2018. Facebook's advertising revenue worldwide from 2009 to 2017 (in million U.S. dollars). *Statista* [online]. Available from: <https://www.statista.com/statistics/271258/facebooks-advertising-revenue-worldwide/> [Accessed 25 October 2018]

Stigler, G. J., 1980. Privacy in Economics and Politics. *Journal of Legal Studies*, (9), 623-644.

Stone, E., and Stone, D., 1990. Privacy in Organizations: Theoretical Issues, Research Findings, and Protection Mechanisms. In *Research in Personnel and Human Resources Management*, 8, 349-411.

Sue, V.M. and Ritter, L.A., 2012. *Conducting online surveys*. New York: Sage.

Taddicken, M., 2014. The 'privacy paradox' in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of Computer-Mediated Communication*, 19 (2), 248-273.

Thornhill, J., 2018. Silicon Valley needs to earthquake-proof its businesses. *FT.com* [online], 9 October 2018. Available from: <https://www.ft.com/content/29c162a2-cb19-11e8-b276-b9069bde0956> [Accessed 10 October 2018]

Tsai, J.Y., Egelman, S., Cranor, L. and Acquisti, A., 2011. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22 (2), 254-268.

Turow, J., Hoofnagle, C.J., Mulligan, D.K. and Good, N., 2007. The federal trade commission and consumer privacy in the coming decade. *ISJLP*, (Issue 3), 723.

United Nations, 1948. The Universal Declaration of Human Rights. *United Nations* [online] 10 Dec 1948. Available from: <http://www.un.org/en/documents/udhr/index.shtml#a12> [Accessed 7 August 2012]

Varian, H., R., 1996. *Economic aspects of personal privacy. Privacy and Self-Regulation in the Information Age, National Telecommunications and Information Administration* [online]. Berkeley: Berkeley University

Varian, H.R., 2009. Economic aspects of personal privacy. *Internet policy and economics*, 101-109.

Varian, H.R., 2014. Beyond big data. *Business Economics*, 49 (1), 27-31.

Vila, T., Greenstadt, R., and Molnar, D., 2004. Why We Can't be Bothered to Read Privacy Policies: Models of Privacy Economics as a Lemons Market. *The Economics of Information Security*, 143–154.

Wakefield, J., 2015. Facebook quizzes; What happens to your data. *BBC News* [online] 26 November 2015. Available from: <https://www.bbc.co.uk/news/technology-34922029> [Accessed on: 25 April 2018]

Warren, S. D., and Brandeis, L. D., 1890. The Right to Privacy. *Harvard Law Review*, Vol. IV, No. 5, 193-220.

WEF, 2017. The Global Risks Report 2017. *World Economic Forum* [online], 11 January 2017. Geneva: WEF. Available from: <https://www.weforum.org/reports/the-global-risks-report-2017> [Accessed 14 February 2017]

Westin, A. F. 1967. *Privacy and Freedom*. New York: Athenaeum.

Westin, A. F., 2003. Social and Political Dimensions of Privacy. *Journal of Social Issues*, 59 (2), 431-453.

US, S.C., 1977. *Whalen v. Roe*. 22 Feb 1977. *United States reports: cases adjudged in the Supreme Court at... and rules announced at... United States Supreme Court*, 429, p.589.

White House, 2012. *Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy*. Washington: The White House.

World Bank, 2003. Information Asymmetry. *The World Bank* [online], September 2003. Available from http://siteresources.worldbank.org/DEC/Resources/84797-1114437274304/Asymmetric_Info_Sep2003.pdf [Accessed 12.11.2018]

Working Party 29, 2014. *Opinion 01/2014 on the application of necessity and proportionality concepts and data protection with the law enforcement sector*.

Article 29 Data Protection Working Party, 536/14/En, WP 211. Brussels: European Commission. Available from:
http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp211_en.pdf [Accessed 07 March 2014]

WSJ, 2012. Facebook Prices IPO at Record Value. *Wall Street Journal* [online], 17 March 2012. Available from:
<http://online.wsj.com/article/SB10001424052702303448404577409923406193162.html> [Accessed 01 March 2013]

Zhu, K., 2002. Information transparency in electronic marketplaces: Why data transparency may hinder the adoption of B2B exchanges. *Electronic markets*, 12 (2), 92-99.

6 Appendices

6.1 Appendix 1: RO1 Survey

This is a link to the survey in relation to Research Objective 1.

https://bournemouthbusiness.az1.qualtrics.com/jfe/form/SV_2hnf0fmj42i6fkx

The figure below is an example screenshot of the survey.

Figure 6-1 RO1 – Example Survey Screenshot

A screenshot of a survey form with a red border. It contains three questions, each with a dropdown menu. The first question asks about home address details, with 'Yes' selected. The second question asks about employment status/history, with 'No' selected. The third question asks about racial or ethnic origin, with 'Not sure/do not know' selected.

Do you believe details that identify your home address details (i.e. postcode/zipcode) are categorized as 'sensitive data' under current EU data protection regulations?

Yes ▼

Do you believe details relating to your employment status/history are categorized as 'sensitive data' under current EU data protection regulations?

No ▼

Do you believe details that identify your racial or ethnic origin are categorized as 'sensitive data' under current EU data protection regulations?

Not sure/do not know ▼

6.2 Appendix 2: RO2 Survey

This is a link to the full survey in relation to Research Objective 2:

https://bournemouthbusiness.az1.qualtrics.com/jfe/form/SV_9nv8LNMTIxeQAtL

The figure below is an example screenshot of the survey.

Figure 6-2 RO2 – Example Survey Screenshot



For each of the questions below, please consider whether you consider personal data (i.e. data about you) to be sensitive or not. If you do not have an opinion either way please select the 'not sure' option.

How do you believe data revealing your racial or ethnic origin should be classified under data protection laws?

How do you believe data revealing your political opinion(s) should be classified under data protection laws?

6.3 Appendix 3: RO3 – EU Experiment

This is a link to the full survey in relation to Research Objective 3 EU Experiment:

https://bournemouthbusiness.az1.qualtrics.com/jfe/form/SV_6tHSzDIZkIUr8QB

The figure below is an example screenshot of the survey.

Figure 6-3 RO3 – EU Experiment Example Screenshot



Please answer the following 15 questions.

Have you ever made up a serious excuse, such as grave illness or death in the family, to get out of doing something?

Have you ever looked at pornographic material?

Have you ever used sex toys?

Have you ever smoked marijuana (i.e., pot, weed)?

6.4 Appendix 4: RO3 – US Experiment

This is a link to the full survey in relation to Research Objective 3 US Experiment:

https://bournemouthbusiness.az1.qualtrics.com/jfe/form/SV_3BOI2xrvxG4ieS9

The figure below is an example screenshot of the survey.

Figure 6-4 RO3 – US Experiment Example Screenshot

Personal Health: Do you have an existing medical or health condition:

- ☐ Yes
- ☐ No
- ☐ Would rather not say

Physical Location: Which of the following best describes the location you are now:

- ☐ At home
- ☐ At work
- ☐ Public place i.e. cafe/restaurant or similar
- ☐ Other
- ☐ Would rather not say



Submit Survey